AUGUST
1958
VOLUME 22
NUMBER 4

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

Edited by Laurance F. Shaffer
Teachers College, Columbia University

Associate Editors: Donald W. Fiske, *University of Chicago.* Harrison G. Gough, *University of California.* Theodore A. Hunt, *Northwestern University.* E. Lowell Kelly, *University of Michigan.* Morrie Krugman, *Board of Education, New York, N. Y.* Bertha M. Luckey, *Board of Education, Cleveland, Ohio.* Fred McKinney, *University of Missouri.* Anne Roe, *New York, N. Y.* Carl R. Rogers, *University of Wisconsin.* Nevitt Sanford, *University of California.* Edward Joseph Shoben, Jr., *Teachers College, Columbia University.* George J. Wyckoff, *University of Pittsburgh.* J. R. Wittenborn, *Rutgers University.*

Arthur C. Hoffman, *Managing Editor*
Helen Orr, *Circulation Manager*
Barbara Cummings, *Editorial Assistant*

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior.

Editorial Office

Manuscripts submitted for *regular* or *early* publication, Brief Reports, and tests for review should be sent to the Editor-elect, Edward S. Bordin, Department of Psychology, University of Michigan, Ann Arbor, Michigan.

Style of Manuscripts. Manuscripts must be prepared in the style described in the *Publication Manual of the American Psychological Association (1957 Revision)*, obtainable for $1.00 from the Washington office of the APA.

Early Publication. Authors may receive the early publication of an acceptable article in the first issue to go to press after its acceptance by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication may be obtained from the Editor or from the Central Office.

American Documentation Institute. To save printing costs, materials supplementary to articles may be deposited with the ADI. To obtain copies, order by Document Number as cited in the reference footnote, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. The indicated fee must be remitted in advance. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Reprints. Authors receive fifty reprints gratis, except for early publication articles and Brief Reports. Additional quantities may be ordered when returning proofs.

Business Office

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must reach the subscription office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $10.00, foreign $8.50, single number $2.50.

Published by the American Psychological Association, Inc.
at Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth St. N.W., Washington, D.C.

Entered as second-class matter, March 31, 1955, at the post office at Lancaster, Pa., under the act of March 3, 1879. Acceptance for mailing at special rate of postage provided for in Section 538, Act of February 28, 1925, authorized May 17, 1949.

© 1958 by the American Psychological Association, Inc.

JOURNAL OF CONSULTING PSYCHOLOGY

August, 1958 Vol. 22, No. 4

Contents

Further Evaluation of the Archimedes Spiral Aftereffect: *Harold Gilberstadt, Jerome D. Schein, and Albert Rosen* - 243

The Spiral Aftereffect in a Neurological Setting: *David Berger, Richard Everson, Louis Rutledge, and Yale David Koskoff* - 249

A Further Study of the Perception of Apparent Motion by Schizophrenics: *Rayford T. Saucer* - 256

Success and Failure in Client-Centered Therapy As a Function of Client Personality Variables: *William L. Kirtner and Desmond S. Cartwright* - - - - - - - - - - 259

The Performance of Psychoanalytic and Client-Centered Therapists in an Initial Interview: *Hans H. Strupp* - 265

A Theoretical Framework for Nondirective Play Therapy: Concepts from Psychoanalysis and Learning Theory: *Dell Lebo* - 275

Personality Patterns of Patients Before and After Application for Psychotherapy: *Norman S. Greenfield* - 280

The Diagnostic Validity of the Ames "Danger Signals": *Elise Elkins* - - - - - - - - 281

An Investigation of Determinants Inducing Individuals to Seek Personal Counseling: *James S. Terwilliger and Fred E. Fiedler* - - - - - - - - - - - - - - - - - - 288

Vocabulary as a Basis for Estimating Intelligence From the Rorschach: *Thomas R. Trier* 289

Extraversion, Neuroticism, Radicalism, and Tendermindedness: *A. W. Bendig* - - - - 292

Handwriting Conventions as Determinants of Human Figure Drawings: *Wayne Dennis* - 293

The Relationship Between Edwards' and McClelland's Measures of Achievement Motivation: *Levon H. Melikian* - 296

Second-Order Factors in Personality Measurement: *Samuel Karson and Kenneth B. Pool* 299

The MMPI Performance of Veterans with Organic and Psychiatric Disabilities: *Joseph J. Motto* - 304

Two New Experimental Scales of the MMPI: *Kenneth B. Little and Jerome Fisher* - - - 305

The Parental Attitudes of Mothers of Schizophrenics: *Marvin Zuckerman, Mary Ollean, and Irwin Monashkin* - 307

Representative Elements in Role Construct Technique: *Spiro B. Mitsos* - - - - - - - 311

The Taylor A Scale and Clinical Measures of Anxiety: *Carl G. Lauterbach* - - - - - - 314

Experience of Clinicians, and Speech Characteristics and Statements of Clients as Variables in Clinical Judgment: *Austin E. Grigg* - 315

A Note on the Reliability of the Leiter International Performance Scale: *Heber C. Sharp* 320

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 75 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style (1). Headings, tables, and references are avoided or, if essential, must be counted in the 75 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 75-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. ——, remitting $—— for microfilm or $—— for photocopies.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual* (1).

Reference

1. American Psychological Association. Council of Editors. *Publication manual of the American Psychological Association (1957 rev.)*. Washington, D. C.: American Psychological Association, 1957.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Further Evaluation of the Archimedes Spiral Aftereffect

Harold Gilberstadt, Jerome D. Schein
VA Hospital, Minneapolis, and University of Minnesota

and Albert Rosen
State College of Washington

The present investigation was stimulated by the report of Price and Deabler (5) who, after testing for the spiral aftereffect on groups of normal, nonorganic psychiatric, and organic brain-damaged hospitalized patients, concluded: "When failure to perceive the aftereffect is found it is almost certain that cortical involvement exists" (5, p. 301). They state that the results of a previous study by Freeman and Josey (1) led them to believe that this phenomenon might be used to develop a technique for the detection of organic brain damage. In a subsequent study, Gallese concluded that "the test almost always indicates organicity when organic scores are obtained although the converse is not true" (2, p. 257).

This study represents an effort to evaluate further the spiral aftereffect as a screening device for cortical brain damage with a VA GM&S hospital sample. In such a setting where large numbers of patients are examined for diagnosis and possible acute and intensive treatment, it would be highly desirable for the clinical psychologist to be able effectively to go beyond the mere assessment of the efficiency of functioning in those areas for which we presently have standardized techniques such as the Wechsler-Bellevue. There is little question that we are still far from able to state with any degree of confidence to what extent specific functional or organic variables contribute to the observed or predicted performance of the patients we examine. Yet it is apparent that our goal in the area of differential diagnosis must be to rely as little as possible on the art of clinical diagnosis which is strong in the tradition of neurology and psychiatry and to continually strive to develop rigorously standardized techniques from which statistically sound predictions of known validity can be generated.

The results obtained by Price and Deabler were of more than usual interest because, considering the complexity of the human organism and the nature of the variables involved in the loss of function from cortical brain damage, it is hard to overcome initially strong scepticism toward the claim that a single, brief test of one modality could develop into an almost infallible sign of such damage. As a consequence, the study by Price and Deabler seemed to demand further evaluation.

Method

Every effort was made to follow the instructions of Price and Deabler for administering the test. An Archimedes Spiral of 920°, or two and one-half turns about its center, was constructed by painting a black spiral on a white background. This wheel was affixed to a reversible motor with a variable speed control which was set at 100 rpm for the experiment. The apparatus was placed in an inside room which was well illuminated by fluorescent lighting. The spiral was placed at eye level and the subject was seated at a distance of eight feet from the apparatus.

One experimenter was employed to test all of the patients. Four trials were given alternating clockwise and counter-clockwise rotations. Scoring was on the basis of 1 or 0 for each trial. Samples of responses to the aftereffect scored 1 were: "It's coming toward me." "Getting larger." "Expanding." "Going in the opposite direction." "Swelling." "Shrinking."

"Opposite." Samples of 0 scoring were: "Nothing." "Stopped." "Standing there." "Standing still." "Same."

A check of scoring reliability by four judges on the first 35 cases showed no disagreements except for one instance in which a judge misread the handwriting on the recording form.

Sample

During a three-month experimental period all available patients admitted to the Psychiatry and Neurology Services respectively were tested. The total number of admissions to Psychiatry were 112 and to Neurology 175.

Table 1

Diagnoses and Scores on Archimedes Spiral Aftereffect of Psychiatric Sample

($N = 87$)

Diagnosis	Distribution of Scores				
	0	1	2	3	4
Organic ($N = 7$)					
Acute brain syndrome associated with alcohol intoxication	1				
Chronic brain syndrome with psychosis	1	1			
Encephalitis, epidemic chronic	1				
Encephalopathy, chronic progressive		1			
Depressive reaction with hypertension and cerebral vascular accident		1			
Alzheimer's disease	1				
Nonpsychotic ($N = 57$)					
Anxiety reaction			2	1	19
Conversion reaction					1
Depressive reaction	2	1	1	2	5
Dissociative reaction				1	2
Obsessive-compulsive reaction					4
Psychophysiologic reaction					4
Adult situational maladjustment					2
Chronic alcoholism					3
Dysocial reaction					1
Emotional instability reaction		1		1	1
Inadequate personality					1
Mental deficiency					1
Psychotic ($N = 23$)					
Involutional melancholia		1			1
Paranoid schizophrenia				3	9
Schizophrenic reaction other		1		2	5
Manic depressive					1

To maintain standard testing conditions, an initial requirement was instituted that only patients who were able to come to the testing room would be included; the criterion being the judgment of the nurse in charge of the ward. As a result of this condition, 16 patients were omitted from the Psychiatry sample. An inspection of the cases showed physical illness, assaultiveness, and suicidal risk to be the bases of the nurses' decisions. Seven patients were omitted through administrative error. Two patients who did come for testing were unable to attend to the task. Table 1 shows the distribution by diagnosis and score of the remaining 87 cases which constitute the Psychiatry sample.

Thirty-five patients admitted to Neurology were omitted from the sample of these cases. Fifteen were discharged before testing could be done, six expired, eleven were bedridden, three were aphasic and after reporting for the test did not have sufficient verbal ability to be tested under the standard conditions. Table 2 gives the distribution by diagnosis and score of the remaining 140 cases.

Mean age of the Psychiatric sample is 41.2 years and mean age of the Neurologic sample 43.4 years. Both groups were tested an average of five days after admission. Diagnoses in all cases were final diagnoses and were uncontaminated by test results which were withheld until all diagnoses had been obtained.

Results and Discussion

In reporting their findings, Price and Deabler used the terms "organic," "brain damaged," and "cortical" as though they were equivalent. That the first two terms may convey very similar meaning is understandable, but the term cortical has a highly specific connotation referring to a particular portion of the brain. Price and Deabler's summary statement led us to conclude that, despite the ambiguous use of these words, their intention was to restrict the test to the detection of *cortical* brain damage. It seemed imperative, then, to dichotomize our organic brain syndrome cases into cortical and noncortical subsamples. Parenthetically, it may be noted that assigning only cortical cases to the category of expected positives was more favorable to the hypotheses of Price and Deabler

Table 2

Diagnoses and Scores on Archimedes Spiral Aftereffect of Neurologic Sample

($N = 140$)

Diagnosis	Distribution of Scores				
	0	1	2	3	4
Organic brain syndrome[a] ($N = 68$)					
Traumatic brain injury	1		1	2	5
Demyelinating disease of brain				1	5
Brain tumor	2			1	1
Cerebral vascular accident	4			3	4
Convulsive disorder	2	1		2	15
Chronic brain syndrome	1		2		1
Meningitis					2
Paralysis agitans		1			3
Other					8
Organic non-brain syndrome[b] ($N = 56$)	9	3	2	8	34
Organic disease not found ($N = 8$)				1	7
Psychoneurosis[c] ($N = 8$)			2	1	5

[a] Specific diagnoses in each category are as follows: *traumatic brain injuries*—2 posttraumatic epilepsy; 1 cerebral concussion, 1 basilar skull fracture, 2 cerebral contusion, 1 penetration wound right parietal area, 1 postconcussion syndrome, 1 posttraumatic encephalopathy; *demyelinating diseases*—3 amyotrophic lateral sclerosis, 1 multiple sclerosis, 1 primary lateral sclerosis, 1 pyramidal tract disease bilateral; *brain tumors*—1 brain tumor, 1 retrobulbar, 1 right temporal, 1 right thalamic; *cerebral vascular accidents*—1 c.v.a., 1 embolus middle cerebral artery with convulsive disorder, 1 occlusion of right internal carotid, 4 cerebral thrombosis, 1 internal carotid artery, 1 post cerebellar artery, 1 old with motor aphasia and right hemiparesis, 1 right hemisphere; *convulsive disorders*—1 black-out spells, 1 cetrencephalic seizures, 6 grand mal epilepsy, 1 grand mal and petit mal epilepsy, 1 grand mal with drug withdrawal, 8 epilepsy of undetermined etiology, 2 psychomotor seizures; *chronic brain syndrome*—1 chronic alcoholism with secondary diffuse cortical atrophy, 1 degenerative brain disease, 2 unknown etiology diffuse; *meningitis*—1 pneumococcal type III, 1 syphilis tertiary; *other*—2 metastatic carcinoma, 1 cervical mass, 1 skull defect right frontal, 1 subarachnoid hemorrhage, 1 tabes dorsalis, 1 headache with retinal hemorrhages, 1 hypertensive cardiovascular disease.

[b] Specific diagnoses are as follows: 1 arachnoid cyst, 1 arteriosclerotic heart disease, 1 back-pain undetermined etiology, 2 Bell's palsy, 1 cervical spine trauma, 1 degenerative cord disease, 1 epididymal cyst, 1 ganglion cyst right wrist, 1 metastatic hypernephroma, 4 low back syndrome, 1 lumbar strain, 1 Meniere's disease, 1 myotonia dystrophica, 1 nephrotosis, 1 posttraumatic neuropathy, 24 herniated nucleus pulposus, 1 occlusion anterior spinal artery, 3 peripheral neuritis, 3 radiculitis, 1 rheumatic heart disease, 1 sacroiliac strain, 1 sciatic syndrome, 1 possible slipped disc, 2 syncope of undetermined etiology.

[c] Specific diagnoses are as follows: 1 anxiety reaction, 1 functional cephalgia, 4 conversion reaction, 1 depressive reaction, 1 psychoneurosis severe with paranoid trends.

than the inclusion of all brain-damage cases would have been.

We relied on the best judgments available to classify our sample on the basis of diagnosis.[1] Nevertheless, we did so with reservation, since a realistic consideration of the problem must lead to the conclusion that within the limits of a study of this type it is not possible to establish a very satisfactory criterion of cortical brain damage. Even with unlimited resources there still would be serious weaknesses in the criterion because of the difficulties inherent in the accurate determination of the localization and extent of cortical brain damage with even the best of current neurological techniques.

Price and Deabler did not elaborate on the basis by which they conclude that their cases of CNS syphilis, cerebral vascular accident, chronic brain syndrome associated with alcoholism, cerebral arteriosclerosis, posttraumatic encephalopathy, and advanced Parkinson's disease have known cortical involvement beyond stating that the diagnoses were based on hospital staff action. Because of our lack of confidence in such diagnoses and for the reason noted above, we were conservative and assumed that only the 30 cases, which included traumatic brain injury, brain tumor, cerebral vascular accident, chronic brain syndrome, and meningitis (listed under the organic brain syndrome subgroup in Table 2), plus the seven cases from the organic subgroup (listed in Table 1) have cortical brain damage of a pathological degree.

Meehl and Rosen's (3) procedure for the evaluation of any assessment or prediction device was followed in the analysis of the test results. In the first three columns of Table 3, f denotes the number of diagnosed cases of cortical damage receiving each score, cf indicates the cumulative frequency and P_1 represents the proportion of the total group of those with cortical damage who scored at or below each scoring interval. The next three columns give the same data for the group without cortical damage. In order to conserve space we have eliminated the intermediate calculations given for illustrative purposes by Meehl and Rosen. H_T represents the proportion of correct classifications for the given cutting line; i.e., the number of valid positives plus the number of valid negatives divided by the total number of cases.

When all cases are considered, it will be noted that the test does no better than the base rates; i.e., if we simply called all cases

[1] We wish to thank E. T. Bell, Emeritus Professor of Pathology, University of Minnesota Medical School, and David Webster of Minneapolis VA Hospital neurology staff for their assistance in categorizing the neurological cases.

Table 3

Cumulative Frequency Tables for Calculation of Optimal Cutting Scores for Archimedes Spiral Aftereffect

	Score	Cortical			Noncortical			
		f	*cf*	P_1	*f*	*cf*	P_2	H_T
Combined sample[a]	0	12	12	.3243	13	13	.0684	.8344
	1	3	15	.4054	9	22	.1158	.8076
	2	3	18	.4865	7	29	.1526	.7896
	3	6	24	.6486	23	52	.2737	.7139
	4	13	37	1.0000	138	190	1.0000	.1600
Psychiatry cases[b]	0	4	4	.5714	2	2	.0250	.9427
	1	3	7	1.0000	4	6	.0750	.9310
	2				3	9	.1125	.8965
	3				10	19	.2375	.7815
	4				61	80	1.0000	.0800
Neurology cases[c]	0	8	8	.2667	11	11	.1000	.7670
	1	0	8	.2667	5	16	.1455	.7311
	2	3	11	.3667	4	20	.1818	.7234
	3	6	17	.5667	13	33	.3000	.6720
	4	13	30	1.0000	77	110	1.0000	.2100

[a] Base rate for cortical brain damage in this sample is .16.
[b] Base rate for cortical brain damage in this sample is .08.
[c] Base rate for cortical brain damage in this sample is .21.

free of cortical damage we would correctly identify 84% of the sample. Using the test, we correctly identify 83.4%, provided that we use a cutting score between 0 and 1. Using the cutting line between 2 and 3 suggested by Price and Deabler, we would classify the cases correctly only 78.9% of the time as compared to the base rates of 84%.

In Psychiatry, where 8% of the sample were diagnosed as having cortical damage, if the cutting line were set between 1 and 2, 93% of the cases would be correctly identified as compared to 92%, using only the base rate and calling all cases non-brain-damaged. Maximum effectiveness is again attained with a cutting line between 0 and 1 where 94.3% of the cases are correctly identified.

In the Neurologic sample, the test is unable to improve upon the base rates. By routinely calling all cases not cortically damaged, 79% would be correctly classified. Using a cutting score between 0 and 1 the test would attain its maximum number of correct identifications, 77%. Use of the test, then, would result in a 2% *decrement* in predictive accuracy.

In addition to the total number of correct classifications, evaluation of a clinical instrument is dependent upon the importance attributed to detecting the greatest number of cortically damaged cases and to the cost and the facilities available for further neurological evaluation of mislabeled, noncortically damaged patients. If one wished to avoid missing any cortical cases regardless of the cost, he would choose a cutting line between 3 and 4 for a population resembling our combined Neurology and Psychiatry samples. This would yield the detection of 64.9% (P_1) of the cortical cases at the expense of misclassifying 27% (P_2) of the noncortical cases. If one were interested in avoiding additional neurological examinations because of staff limitations, he would select a cutting line between 0 and 1 where 32.4% of the cortical cases are identified, and only 6.8% of the noncortical cases are misdiagnosed by the test.

Table 4 contains a similar analysis of the data presented by Price and Deabler (5), Gallese (2), and Page, Rakita, Kaplan, and Smith (4). We have also repeated the data for our combined Neurology and Psychiatry samples to simplify comparison of all four studies. Since none of the other three studies

report the base rates for their populations, we have used the empirical rate obtained from our combined sample (.16) as the estimate with which to make the necessary calculations.

It will be noted in Table 4 that the maximum total hit rate (H_T) varies from 96.4% in the Price and Deabler study, to 90.4% for Gallese, to 81.0% for Page *et al.*, to 83.4% for the Minneapolis sample, and that the cutting line at which the maximum is obtained alternates from a point between 2 and 3 in the first two studies to between 0 and 1 for the last two studies. Gallese stated that his method was more apt to reduce the proportion of false positives inasmuch as he continued to press for a "normal" response if it were not at first obtained. He was also more liberal in scoring responses in the "normal" direction. In the Page *et al.* study, the spiral was revolved in one direction only for three trials rather than in two directions for four trials.

The present study, however, does not differ from Price and Deabler's in test administration. The reduction in total hit rate to 83.4% (using a cutting line between 0 and 1) is in our opinion largely a result of differences in sampling. Our sample was drawn randomly. All consecutive admissions who could be tested were tested, with exceptions as noted above. Furthermore, our patients were probably less deteriorated than those selected by Price and Deabler though we cannot be certain because of the paucity of information they provided. We feel that the lack of description of their sample represents a very serious deficiency in the usefulness of their results. It is possible to select samples of brain-damaged and psychotic populations such that the brain-damaged, being severely deteriorated, can fail almost any task given them and the psychotics, being relatively intact, obtain a high rate of success. Such a procedure yields a "mythical population," i.e., one which does not occur in practice. Tests validated on such groups will give deceptively good discrimination between groups, only to fail on cross-validation.

It is perhaps well to note that developing a mythical population is not entirely without virtue. If the means of obtaining such a sample are carefully reported, it may be possible to locate or duplicate a reasonably similar sample. On the other hand, nonrandomly selecting a sample may result in one being unable to validly predict what the results might be even in one's own institutional setting where the sample was drawn. We can only speculate, therefore, as to the nature of the differences between the other samples and the one drawn for this study. Gallese, for example, reports that the median duration of hospitalization for his organic group is 65 months and for his schizophrenic group 33 months in contrast to our groups who were tested an average of five days after admission to the hospital. There are also substantial differences in the percentage of cases of CNS syphilis in our samples and those of Gallese, and it is probable that this would be the case in a comparison with the Price and Deabler sample. The fact that Gallese obtains results intermediate between ours and those of Price and Deabler might suggest that he has drawn a sample which more closely approximates theirs on such significant variables.

Table 4

Percentage of Hits for Each Cutting Line in Different Hospital Samples Assuming a Base Rate of .16

	Percentage of Hits in Each Study			
Cutting Lines Between	Price & Deabler	Gallese	Page *et al.*	Minneapolis
0 and 1				
H_P (Positive hit rate)	100.0	*	43.2	47.4
H_N (Negative hit rate)	92.9	*	91.8	87.9
H_T (Total hit rate)	93.6	*	81.0	83.4
1 and 2				
H_P	91.5	*	*	40.0
H_N	94.7	*	*	88.6
H_T	94.3	*	*	80.8
2 and 3				
H_P	87.4	72.0	*	37.8
H_N	98.2	93.6	*	89.6
H_T	96.4	90.4	*	79.0
3 and 4				
H_P	78.9	*	*	31.1
H_N	99.7	*	*	91.6
H_T	95.5	*	*	71.4

* Data unavailable.

Table 4 also contains for each cutting line the positive hit rate

$$\left(H_P = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}\right)$$

and the negative hit rate

$$\left(H_N = \frac{\text{true negatives}}{\text{true negatives} + \text{false negatives}}\right).$$

Thus, in Price and Deabler's study, 100% of cases obtaining a score of 0 are diagnosed as cortically damaged, while this is true for only 43.2% of the Page *et al.* sample and 47.4% of our sample. Conversely, if we wished to maximize the negative hit rate, the cutting line should be placed between 3 and 4 where the rate of correct negative predictions was 99.7% for the Price and Deabler sample and 91.6% in our sample. One can, then, select a cutting line depending on (*a*) which sample most closely resembles the one in which the test is to be used and (*b*) whether one wishes to maximize H_P, H_N, or H_T.

In most clinics, use of the spiral aftereffect as a screening device will be considered in relation to the efficiency of the instrument as well as the kinds of decisions which will be made on the basis of positive results. Assuming that further cross-validation will not cause much shrinkage, it might be practical to use the test with samples which resemble our Psychiatric sample where only 8% false positives were obtained with a cutting score between 0 and 1. Considering the small amount of time and effort involved, it would be feasible to administer the test to all patients admitted to Psychiatry and to recommend a neurological workup for those who score below this cutting line.

The test would not seem to have utility for the screening of Neurology patients. Since there is such an unfavorable ratio between valid and false positives it would not appear feasible even to advise the neurologist to use the test findings as an added weight in his diagnostic formulation.

Summary

Evaluation of 87 consecutive testable admissions to the Psychiatry Service and 140 to the Neurology Service of the Minneapolis Veterans Administration Hospital by means of the Archimedes Spiral Aftereffect indicated that the test had very limited clinical usefulness for detecting pathological cortical brain damage. Use of the test did not significantly improve the diagnostic efficiency which would be obtained by using the base rates.

Problems of sampling and criteria were discussed. Differences between this study and those previously reported appear to be largely a function of differences in the samples studied. Problems inherent in the use of nonrandom samples for purposes of validation were also pointed out.

Received August 21, 1957.

References

1. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, **62**, 794–796.
2. Gallese, A. J. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254–258.
3. Meehl, P. E., & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.
4. Page, H. A., Rakita, G., Kaplan, H. K., & Smith, N. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, **21**, 89–91.
5. Price, A. C., & Deabler, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Spiral Aftereffect in a Neurological Setting

David Berger,[1] Richard Everson, Louis Rutledge[2]
VA Hospital, Leech Farm Road, Pittsburgh
and Yale David Koskoff[3]
Montefiore Hospital, Pittsburgh

The spiral aftereffect may normally be seen when a circular disk supporting an Archimedes spiral is rotated within certain limits of speed. During the period of rotation, the spiral appears either to expand or to contract, depending on both the direction of rotation and the direction of the coil of the spiral. When the rotating movement of the disk is interrupted, an aftereffect occurs so that the spiral appears to rotate in the opposite direction.

Like many techniques being reactivated in current perception research, the spiral aftereffect has a substantial history. In 1850, Plateau, the Belgian scientist, reported the first systematic study of the moving spiral. More lately, the relation between brightness, direction, velocity, etc., and the perception of the aftereffect has been investigated (3, 9).

Along with the interest in the various parameters of the experience is the attempt to relate the performance on the spiral to discrete mental functions and to brain damage. Freeman and Josey's paper (1) on the correspondence between memory impairment and reduced ability to perceive the aftereffect, spurred Price and Deabler (7) to measure the relationship between the aftereffect and cortical damage. The results of this latter effort are most impressive in that virtually no overlap in performance was found between a group of brain-damaged patients and two control groups. In the brain-damaged group 98% failed to perceive the aftereffect in each of the four consecutive trials, while 95% of the nonorganic psychiatric population, and 92.5% of the normal population, were able to perform satisfactorily on each of the four trials. In a subsequent paper, Garrett, Price, and Deabler (2) repeated their original work with 40 chronic brain-injured patients. Their findings substantiate their previous results, inasmuch as 97.5% of the sample failed to perceive the aftereffect in each of the four trials. Page *et al.* (6) have provided another test relating performance on the spiral to cortical damage. While their results were in the same direction as Price and Deabler's, the ability of the spiral test to discriminate was patently less efficient. In another study conducted by Smith and Goldberg,[4] the spiral was administered to normals, psychiatric patients, organics, and postshock patients. The test proved capable of discriminating between pathological processes and normals, but apparently failed to distinguish organic impairment from the functional disorders.

The present study was undertaken to cross-validate the findings reported by Price and Deabler and particularly the statement, "While adequate explanation of the phenomenon remains to be hypothesized and verified, the results are such as to justify present clinical use of this technique in diagnosis of organicity (7, p. 302). In addition, the following problems were evaluated: (*a*) the effect of the direction of the coil of the spiral on difficulty and duration of the after-

[1] Now at Crile VA Hospital, Medical Service.

[2] Now at VA Hospital, Denver.

[3] The authors wish to express their appreciation to Joen Fagan, graduate student at Pennsylvania State University, and Ted Rice, graduate student at Western Reserve University, who assisted in collecting and evaluating the data.

[4] P. Smith and L. Goldberg. Personal communication. April, 1957.

effect, (*b*) the relation between visual acuity and the aftereffect, and (*c*) the relation between the aftereffect and various neurological indicia.

Method

Subjects

With the exception of bedfast patients and those patients who were too incapacitated to cooperate, all patients hospitalized during a three-month period on a neurological ward in a VA hospital were selected for study. Some months after the experimental data were collected, the patients' clinical charts were reviewed for information relevant to the following areas: diagnoses,[5] admission symptoms, electroencephalographic (EEG) reports, pneumoencephalographic studies, skull X-ray reports, findings on spinal fluid examination, and reports on visual fields. In certain cases, the chart failed to yield data for each variable. The availability of this type of information was largely a function of each patient's condition and the consequent need for laboratory study.

Apparatus

A six-inch Archimedes spiral formed in 2½ convolutions was drawn and photographed. Two identical spirals were developed, except that the second was the photographic reversal of the first. The two spirals were then mounted on two 6-in., circular, metal disks. A small metal collar was centered and fastened to the reverse side of the disks. The collar was threaded to complement a machine screw which had been fitted to the center post of a phonograph turntable. The turntable, together with its driving mechanism, was supported on an adjustable laboratory stand so that the target spiral could be presented at eye level to a subject (*S*) seated at a distance of 8 ft. The turntable was activated by a 78 rpm unidirectional motor. Spiral A seemed to expand while rotating and normally would be said to contract in the motionless aftereffect condition. Spiral B, since it was the photographic reversal, and since the motor was unidirectional, appeared to be contracting while rotating and to be expanding in the motionless aftereffect condition.

[5] All diagnoses reported in this study were established after weeks of observation and study.

In addition to the spiral apparatus, a Snellen chart was used to assess visual acuity. As with most work on visual acuity, the minimal separable or visual angle was considered the invariant. However, since the influence of distance is not clearly defined in present-day knowledge of visual acuity, it was also controlled. The size of the room used for testing was such that the *S*s had to read the chart from a distance of 18 ft. By convention, readings of visual acuity, thus obtained, would be adjusted to comply with a 20-ft. distance from the chart. Since the distance error in the ratio was constant in this study, it was not necessary to make this adjustment. In all calculations based on Snellen chart readings, the denominator in the ratio was used exclusively (i.e., with readings such as 18/30 or 18/20, the computations were performed only on the denominators).

Testing Procedure

Two experimenters (*E*s) were present during all testing. One *E* typically discussed the procedure with *S* and recorded behavior. The second *E* observed and adjusted the apparatus. The *S*s were encouraged to wear glasses. When glasses were used, they were required throughout the testing. All visual tasks in this experimentation were viewed binocularly under artificial overhead light.

The patient was told that the tests about to be administered were a special eye examination. The Snellen chart was then called to the patient's attention. He was told to read each succeeding line from left to right until interrupted. The score on this task was the denominator of the acuity ratio of the most difficult line read faultlessly. The criterion required that the line be read perfectly twice and that the succeeding line be faulted twice. The latter condition was obviously inapplicable in the case of an errorless performance on the entire chart. The *S*s were prompted to rest their eyes whenever a line was failed.

The spiral task was then introduced. The two spirals were presented twice each for every *S* in "controlled" random order for a total of four exposures per *S*. The order of presentation was thus randonized for blocks of six patients: ABAB; ABBA; AABB; BBAA; BABA; BAAB. Hereafter, each of the four exposures are referred to as a presentation. Between presentations, a cardboard shield was used to cover the apparatus and the *S* was instructed to close his eyes to rest them. Every effort was made to make the interval between presentations uniform both for time and auditory cues. This was necessary

in order to reduce the extraneous information which might be used to anticipate which spiral was being readied for the next presentation.

The instructions for the spiral task were, "Now please look at the center of the design on the wheel and keep looking right at the center." At that point, the motor which activated the spiral was switched on. After 10 seconds, the *S* was asked to describe what he saw. All *S*s included in the experimental sample were able to describe some form of expanding or contracting experience while the spiral was rotating. The form described depended, of course, on which of the two spirals was being used during the initial presentation. After thirty seconds of observation, the power was cut and the spiral came to rest. The average temporal lag between the time the power was cut and the time the spiral came to a complete stop was 2½ seconds. During this interval, *S* was cautioned again to "Keep your eyes on the center of the design on the wheel." After five seconds elapsed the *S* was asked a series of questions in the following sequence: (*a*) What do you see now? (*b*) Does it look different now? (*c*) Does it look like it's moving now? (*d*) Does it look like it's getting larger or smaller now?

The procedure was in marked contrast to the one employed by Price and Deabler. They assert "there was little ambiguity as to the presence or absence of the aftereffect with most *S*s" (7, p. 302). In the pretesting of the apparatus used in the present study, there were many occasions when *S*'s verbal report failed to establish the legitimacy of the aftereffect experience. The graduated series of questions was designed to uncover the essence of *S*'s perceptual experience. A response at any level of questioning which indicated that the *S* experienced a change either in direction, size, or in distance was acceptable as long as the experience was in the expected direction (i.e., with a spiral which would normally be seen as contracting in the aftereffect condition, reports such as moving away, increasing distance, or decreasing size were all equally "correct"). All acceptable responses were scored plus. When *S*'s response for any presentation was "correct" at any one of the four levels of questioning, that presentation was completed. Thus, all four levels of questioning were used only for responses which were not sufficiently clear to be scored with confidence. A response which ultimately failed to convince the *E* that the *S* perceived the aftereffect or which, at any level of questioning, described the movement in the "wrong" direction, was scored as a minus performance for that presentation. Allowedly, the verbiage employed in the questions tended to promote ideation about movement, distance, and size. Nevertheless, a "correct" response had to incorporate direction of change. In spite of the nature of the questions, the direction of change had a chance expectancy of two correct guesses in four presentations. Accordingly, the *S* was classified as a "Seer" if he scored plus in at least three of the four presentations. *S*s earning two or more minus scores were classified as "Nonseers."

In the case of "Nonseers," the fourth presentation completed their performance. The "Seers," however, were asked to continue. The particular spiral used in the fourth presentation was retained through four additional trials, during which the duration of the aftereffect was ascertained. These four duration exposures will hereafter be referred to as trials. The time per trial was recorded in seconds from the instant the spiral was set into motion, through the 30-second exposure period, and up to the moment the *S* indicated that the aftereffect seemed to be extinguished. The directions for these four trials were, "Do you remember just a moment ago when we stopped the wheel, you told me that the design on the wheel looked like it was '_____' (the subject's own descriptive phrase was used). Now when we start the wheel this time, look at the center again and keep looking at it after I stop it. Then tell me when the '_____' (inserting the same phrase again) first stops." After completing four trials, the "Seers" were excused from the test situation.

Results

The final sample tested numbered 110 patients, all of whom were admitted to the hospital ostensibly for the diagnosis and treatment of neurological conditions. The mean age of the sample was 38.4 years, with an *SD* of 10.99 years and a range of 21 to 64. The primary diagnoses represented in the sample are presented in Table 1 in order to provide some notion of the character of the ward population at the time the testing was conducted. Of the 110 patients tested, 31 (28%) were classified by their performance as "Nonseers." The incidence of "Nonseers" for the various diagnoses is also presented in Table 1.

Comparisons between "Seers" and "Nonseers"

The means of the ages for the "Seers" and "Nonseers" were contrasted, and the lack of significance is summarized in Table 2. In order to compare the means of the Snellen chart performances of the "Seers" and "Nonseers," a square root transformation was used since the distributions were rather severely skewed. The results of this comparison as presented in Table 2 were significant ($< .01$). From this finding it may be said that "Nonseers" have less acute vision than "Seers" and that this relationship does not appear to be related to age.

Comparison of the Two Spirals

The distributions of duration times for each of the two spirals also required normalizing.

Table 1

Diagnoses and Incidence of "Nonseers"

Diagnosis	N	"Non-seers"	Diagnosis	N	"Non-seers"
Multiple sclerosis	24	14	Syncope	3	0
Encephalopathy	15	2	Arteriosclerosis	3	1
Epilepsy, grand mal	15	2	Intracranial aneurysm	3	1
Parkinsonism	7	4	Cerebral thrombosis	3	0
Neurotic reaction	7	0	Alcoholism, chronic	2	0
Epilepsy,			Cephalgia	2	0
Jacksonian	4	3	Amyotrophic lateral		
psychomotor	4	0	sclerosis	2	0
idiopathic	3	0			

Note.—There was also one each of the following diagnoses: Muscular dystrophy, migraine, toxic neuropathy,* papillitus optic nerve, cerebral embolism,* Friedrich's ataxia, poliomyelitis, progressive spinal muscular atrophy, neuralgia, Huntington's chorea,* osteoarthritis, transverse myelitis, and schizophrenic reaction, paranoid type.*

* Denotes a "Nonseer."

Again the square root transformation was employed. The results of the comparison of the mean duration times obtained on the two spirals are presented in Table 3. It is to be remembered that the duration trials were administered only to "Seers." Accordingly, Spiral B, which normally is seen as expanding in the motionless condition, is seen for a significantly longer time than is Spiral A ($< .001$).

Next, the relative difficulty of the two spirals was examined. Spiral A (contracting aftereffect) yielded 60 failures (27%) out of 220 exposures. Spiral B yielded 36 failures (16%) in a similar number of exposures. These percentages were tested by the critical ratio formula and proved quite significant ($< .01$). Thus Spiral B (expanding aftereffect) was appreciably less difficult than Spiral A. The performance on the two spirals was then distinguishable both for duration and difficulty of the aftereffect. The aftereffect on Spiral B was easier to see and was seen for a longer time interval.

Relation Between Spiral and Clinical Criteria

The information gathered from the patients' clinical charts was related to the performance on the spiral aftereffect. This approach involved analyzing the chart data on five independent measures, as well as via a global impression of brain damage. The five indices used were EEGs, pneumoencephalograms, skull X rays, spinal fluid, and visual fields. The global judgments integrated these five indices, as well as the admission symptoms and the diagnoses into a single, aggregate judgment.

Two judges independently evaluated the abstracted chart summaries for each of the five indices with 20 of the patients. These 20 charts were selected in a random manner with the one stipulation that information on the five indices was available. The judges were in-

Table 2

Age and Visual Acuity in "Seers" and "Nonseers"

Variable	Mean scores		*SD*		*t*	*p*
	"Seers"	"Non-seers"	"Seers"	"Non-seers"		
Age	37.71	40.26	10.41	12.35	1.00	—
Visual acuity[a]	4.4871	5.9007	1.3021	2.4707	2.98	$<.01$

[a] Normalized via a square root transformation.

Table 3

Comparison of Duration Time[a] on Spirals A and B

Spiral	*N*	Mean	*SD*	*t*	*p*
A	152	6.3861	.6001	6.2969	<.001
B	163[b]	6.8879	.8018		

[a] Normalized via a square root transformation.
[b] One patient failed to perceive the aftereffect on one of the duration trials on Spiral B.

structed to assign each case to one of the three categories: obvious brain damage, questionable brain damage, or no evidence of brain damage for each variable (EEG, etc.). The two judges agreed perfectly on 84% of the ratings. With this degree of consistency, it was possible to use the judgments of a single judge for the purpose of relating the five indices to the spiral performance. Whenever data were not available, the patient was excluded from the analysis on that particular index. The contingency coefficients between the spiral performance and the neurological measures, as well as the levels of significance, are presented in Table 4. Only chance association was demonstrated between the performance on the spiral test and the three clinical variables: EEGs, pneumoencephalograms, and skull X rays. However, in the case of the visual field examinations and the spiral fluid, significant correlations (both $< .05$) were obtained with the performance in the aftereffect situation. The global estimate of cortical damage was made by one of the authors.[6] The contingency coefficients along with the level of significance ($< .05$) are also presented in Table 4.

Discussion

When certain stimulus patterns are observed in motion at critical speeds, the activity may appear to continue for some appreciable time after the objective motion ceases. Such experience is often termed motion afterimage. The spiral aftereffect is a form of subjective motion afterimage, complicated by a negative experience of changing size or depth. As such, it incorporates radial movement, apparent movement toward or away from an observer, and gamma movement, another form of apparent movement usually associated with a swelling of the size of a suddenly illuminated stimulus.

The Spiral Aftereffect as a Clinical Tool

The results of the present study tend to confirm the general usefulness of the spiral aftereffect as a heuristic laboratory technique for studying brain pathology. There is, however, some real question about the propriety of the technique in clinical practice. The spiral task as explored in the present study does not appear to live up to the claims made for it by Price and Deabler as an individual diagnostic measure.

The critical need in clinical practice is for a technique sufficiently valid to be utilized with ambiguous and difficult problems in the diagnosis of brain damage. When performance on the spiral was compared with patients' diagnosis in the present study, the efficacy of the aftereffect as an individual diagnostic instrument seemed dubious. The disparities between the results in the studies by Price and Deabler and the present work may be attributable to many factors. For one thing, the degree of obviousness of brain dysfunction in the organic samples tested may have been at variance. In addition, some very real differences in methodology were employed. The scoring scheme in the present study made no distinction between verbal reports of changes in size and depth. In the study by Price and Deabler, size changes were given greater recognition in scoring. The justification for each specialization in scoring was not apparent to the present authors and, consequently, any response which indicated

Table 4

Relationship Between Performance on the Spiral Task and Clinical Neurological Tests

Variable	Contingency coefficient	Significance
EEG	.11	—
Pneumoencephalogram	.03	—
Skull X ray	.05	—
Spinal fluid	.24	.05
Visual fields	.27	.05
Global analysis	.25	.05

[6] These judgments were made by the neurologist cooperating on this project.

an appreciation of negative perceptual experience was fully accredited. In addition, the intention in the present study was to provide *S* with every chance to report an experience which may normally appear somewhat freakish and unnatural. Such an experience, it was reasoned, might be rather difficult to verbalize. Frequently it was found that *S*s did quite well on the task only after they fully understood the inquiry. At times this required the complete battery of four questions in order to elicit a satisfactory report. It is suggested that some of the inability of the organically impaired to report the aftereffect in Price and Deabler's paper may have been in part a consequence of communication failure rather than perception failure.

The visual acuity findings are further reason for doubting the legitimacy of the spiral aftereffect as a clinical measure of brain damage. The need to consider visual acuity in a study of perceptual processes is discussed by Yates (10) who admonishes against the omission of a measure of visual acuity in the development of perceptual tests of brain damage. This is an especially pertinent recommendation in instances where the experimental sample includes disease processes such as multiple sclerosis, wherein repeated atrophy of the optic nerve may frequently occur (4, 8). When atrophy develops, it seems to be present most often in the area of the maculopapillar bundle, which is involved in central vision. Since the spiral task involves foveal vision, in a large measure, the need for an index of visual acuity is apparent.

In a superficial attempt to explore further the relationship between visual acuity and brain damage, the Snellen chart readings were contrasted in the three groups differentiated in the global neurological appraisal. No relationship was found between these estimates of brain damage and visual acuity. The question as to the nature of the relationship between visual acuity and performance on the spiral must come under closer scrutiny before the technique can be used with confidence in clinical practice.

The Comparison of the Two Spirals

The performance on the two spirals indicates that the level of difficulty in perceiving the aftereffect, as well as the length of the aftereffect, was quite distinct. One of the best documented facts about the perseveration of movement afterimage is its dependence upon the velocity of the revolving stimulus. However, this factor was controlled in the present study, except for changes in line voltage which should have affected both spirals without bias. The finding that a spiral which seems to be expanding has a longer period of after-activity is consistent with observations of duration in gamma movement studies. In the latter studies, the expansion phenomenon, which occurs when a light source is introduced in a visual field, has been contrasted with the shrinkage which is reported when the light is dimmed or extinguished. As in the observations on the two spirals, the spreading movement occupies a significantly longer time interval than the shrinking movement. The two spirals were also compared to determine if they were equally effective in differentiating cortical damage as determined by the global clinical judgments. In spite of the fact that Spiral A elicited a greater absolute number of failures, there was no difference in efficiency of the two spirals in detecting brain damage.

The Spiral Aftereffect and Vision

The degree of relationship between the spiral performance and clinical visual tests inspired conjecture about the possible connection between deficit on the spiral task and injury to areas of the optic apparatus. The fact that the task was mediated through visual structures added to the temptation to localize the brain damage. Accordingly, all the patients whose clinical charts suggested a global evaluation of obvious or questionable brain damage were re-evaluated in an effort to estimate the localization of the presumed damage. However, the level of reliability for these judgments proved unsatisfactory, and the effort was abandoned for the present paper.

The evidence in this study suggested rather pointedly that two somewhat antagonistic aspects of visual functioning were simultaneously impaired in "Nonseers." On the one hand the ability of the "Nonseers" to make fine visual discriminations (visual acuity) was notably lessened. Such a function, according to various theories of vision, should

require a high degree of insulation among juxtaposed individual pathways as well as a great number and superior differentiation of retinal receptors. The aftereffect task, on the other hand, would require either centrally or in the peripheral system rather complex neural interaction. Thus the dual failure of "Non-seers" on visual acuity and spiral aftereffect tasks provides concurrent behavioral evidence of deficit in anatomical insulation and physiological interaction.

Conclusions

The purpose of this study was to evaluate the use of the spiral aftereffect as an index of brain damage in a neurological setting. In addition, certain neurological specifications of organic impairment were employed as criteria against which perceptual experience was measured. And, finally, the direction of the coil of the spiral was varied, and the visual acuity of the subjects was tested in order to evaluate their effect upon performance. The results are as follows:

1. Of a sample of 110 patients admitted to a neurological ward, 28% were unable to perform satisfactorily on the spiral task.

2. Those who performed satisfactorily on the spiral task had significantly better visual acuity as measured by Snellen chart performance, but were not distinguishable in age from those whose performance on the spiral task was considered unsatisfactory.

3. Whether the aftereffect took the form of expansion or contraction was important for the ease with which it was reported, as well as for the length of time it was subjectively perceived.

4. Performance on the spiral was related to spinal fluid examinations, visual field studies, and global judgments of the likelihood of brain damage. Performance on the spiral was not related to EEG reports, pneumoencephalographic studies, or skull X-ray reports.

The degree of relationship demonstrated between estimated brain damage and spiral performance was taken as vindicating further exploration of the technique. It was suggested, however, that caution be exercised in the clinical use of the test, particularly in those situations in which the demand for such a test would be the most imperative, i.e., in the evaluation of individual cases where the diagnosis is less than straight-forward and direct.

Received August 5, 1957.

References

1. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, **62**, 794–796.
2. Garrett, E. S., Price, A. C., & Deabler, H. L. Diagnostic testing for cortical brain impairment. *Arch. Neurol. Psychiat.*, 1957, **77**, 223–225.
3. Gates, L. W. The aftereffect of visually observed movement. *Amer. J. Psychol.*, 1934, **46**, 34–46.
4. Grinker, R. R., & Bucy, P. C. *Neurology*. Springfield, Ill.: Charles C Thomas, 1949.
5. Kohler, W., & Wallach, H. Figural aftereffect: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.
6. Page, H. A., Rakita, G., Kaplan, H. K., & Smith, N. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, **21**, 89–91.
7. Price, A. C., & Deabler, H. L. Diagnosis of organicity by means of the spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.
8. Schumacher, G. A. Multiple sclerosis and its treatment. *J. Amer. med. Ass.*, 1950, **143**, 1059–1065.
9. Wohlegemuth, A. On the aftereffect of seen movement. *Brit. J. Psychol., Monogr. Suppl.*, 1911, **1**, 1–117.
10. Yates, A. J. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, **41**, 359–379.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

A Further Study of the Perception of Apparent Motion by Schizophrenics

Rayford T. Saucer
VA Hospital, Perry Point, Maryland

The perception of visual apparent movement in normal subjects has been widely studied. It is known that when two lights are exposed alternately at a low rate of alternation, the perceptual emergent will be that of two lights flickering alternately. If the rate of alternation is raised, the subject will report a single light which swings smoothly back and forth. This state is known as optimal movement or beta movement (5). If the frequency of alternation is again raised, the subject will report seeing two lights which flicker simultaneously. It has also been noted that at the boundary between the disappearance of beta motion and the perception of simultaneous flicker, a formless dark object in motion may be perceived, called omega movement (4).

Research (1, 2, 3, 7) indicated that one of the most common results of organic damage to the cortex is the loss or impairment of the ability to perceive visual apparent movement. Saucer and Deabler (5), in an experiment designed to explore the use of beta movement as a diagnostic procedure in this area, used the appearance of omega movement as an index of failure to perceive optimal movement. They found that normal or control subjects were able to perceive beta movement at a much higher rate than organics.

They also found that a group of undifferentiated schizophrenics showed some degree of impairment in the ability to perceive optimal movement, although a group of lobotomized schizophrenics showed no measurable deficit. The present paper is a report of an experiment designed to further explore the loss of ability to perceive optimal movement in schizophrenic individuals.

The apparatus used in this study was a low frequency, square-wave generator, with a Hayden motor sweeping from 2.5 to 10 cycles per second in three minutes. The square-wave generator gated two Sylvania R1130B glow modulator tubes 180° out of phase. A green "Dialite" fresnel lens (No. 34–841) was placed over the end of the R1130B tube. The fresnel lenses were 2.5 inches center to center and 8 inches behind a ⅛-inch opal Lucite screen 12 × 12 inches square. The intensity of the lights was adjusted to optimal movement for all subjects at 2.5 cycles per second and a mean breakdown rate of beta movement for a pilot group of controls of about 5 cycles per second.

Subjects and Method

Subjects (*Ss*) were 21 undifferentiated schizophrenics, 21 paranoid schizophrenics, and 10 control *Ss*. The patients were drawn from the NP population of Perry Point Veterans Administration Hospital, while the control group was composed of hospital nursing assistants and other nonprofessional hospital personnel. Patients were selected by ward psychiatrists and their diagnosis made known after they were tested.

The *Ss* were brought one at a time into the testing room and asked if they would consent to sitting in a dark room for a few minutes. Those who were mute or incooperative were released immediately. The remaining *Ss* were given three minutes of dark adaptation, during which time the experimenter engaged them in trivial conversation.

At the end of the three minutes, the stimulus lights were turned on at an alternation frequency of 2.5 cycles per second and the *S* asked to tell the experimenter what he saw. Almost invariably the *S* reported seeing a single light moving back and forth across the screen. The sweep drive was then turned on and the *S* asked to report any changes in the light. Most *Ss* immediately reported "it's getting faster," but a simple "yes" seemed to be an adequate explanation.

When the *S* reported a dark object in motion, the sweep drive was stopped and the frequency in cycles per second read from a calibrated dial. If the *S* did not report a single light moving back and forth which was replaced by a dark object in motion, no score was given.

Results

The rate at which omega movement appeared was taken as the rate at which optimal movement had ceased. The mean rates and their associated standard deviations are given in Table 1.

The results are presented in Table 2 in terms of t and the Mann-Whitney U statistic (6, pp. 116–127). Although the differences between the schizophrenic group and the paranoid and control groups are statistically significant on the basis of the computed t, nonparametric analysis was also carried out because (*a*) it is evident that the variance is correlated with the mean, and (*b*) nothing is known concerning the population distribution of omega thresholds. The Mann-Whitney x/sigma values are equal to critical ratios when N is sufficiently large and hence may be interpreted by reference to a table of

Table 1

Threshold Means and Standard Deviations of Groups

Group	Mean	*SD*
Schizophrenic	3.428	.766
Control	5.15	1.39
Paranoid	5.417	1.96

x/sigma for a normal distribution. It can be seen that the values are sufficiently large to allow the null hypothesis of no true difference between the schizophrenic group and the control and paranoid groups to be rejected at a comfortable level of significance.

Since the statistical analysis does not clearly show the degree of overlapping between the different categories, an analysis by approximate quartiles is given in Table 3. It can be seen from this table that whereas the majority of the schizophrenics are below the cutting score of 4.130 cycles per second, the larger part of the paranoid *Ss* tested fall into the higher quartile ranges. If we wager that all scores below the median belong to schizophrenics and all high scores to paranoids, then we shall be correct in 33 cases out of 42. If we make the same wager for the normal-schizophrenic grouping, we shall be correct in 25 cases out of 31. The associated percentages of correct prediction are 78.5% for the schizophrenic-paranoid group and 80% for the schizophrenic-normal group where 50% correct predictions can be made by change alone.

Table 2

Significance of Mean and Rank Differences

Groups compared	t	Significance	x/sigma	Significance
Schizophrenic-Control	3.67	.001	3.80	.00009
Paranoid-Control	NS		NS	
Schizophrenic-Paranoid	4.35	.001	3.91	.00005

Discussion

The primary hypothesis of a difference in apparent motion perception between normal and undifferentiated schizophrenics seems to be tenable. The fact that these differences are more marked in this experiment than in the earlier experiment (5) may be due to the increase in sample size. The combined findings from the two experiments suggest that there is a noticeable deficit in the perception of optimal movement by the schizophrenic.

Since the results of the experimental procedure do agree reasonably well with the presence or absence of schizophrenia as established by other diagnostic criteria, the use of apparent motion perception as an added diagnostic procedure seems warranted. At the same time, due to the use of relatively complex electronic test equipment, the method seems more useful in connection with research procedures than as a routine diagnostic test.

The evidence concerning the "normal" perception of apparent movement in the paranoid schizophrenic is of considerable interest and

Table 3

Distribution by Cutting Scores

Cutting score	Schizophrenics	Paranoids	Controls	N
2.50–3.375	10	4	0	14
3.38–4.125	8	2	3	13
4.130–4.875	3	5	2	10
4.88–12.00	0	10	5	15
Total	21	21	10	52

merits separate discussion. The differences between schizophrenics and paranoid schizophrenics are so great as to suggest the presence of an entirely different pathology. It has been argued earlier (5) that the ability to perceive apparent movement is an index of cortical efficiency. If this is so, then the paranoid schizophrenic shows little evidence of lowered cortical efficiency in spite of the fact that as a group the paranoid schizophrenics tested showed considerable behavioral pathology.

In a first approximation to a theory of paranoid schizophrenia, it may be assumed that there are certain individuals who are genetically equipped with nervous systems which are highly efficient in the organization of perceptual processes. This assumption is not entirely groundless, since it is also implied in a genetic theory of intelligence.

Such an individual might be expected to go beyond the norm in organizing his perceptions into a coherent system. If there are severe personality problems present, this structure may become contaminated by personal reference or bias. Since the bias introduces some distortion of reality, it may be expected to change the individual's behavior towards the external world. This in turn begins to produce reactions on the part of society which must be incorporated into the meaningful if paranoid structure of perceptual organizations. The entire process may become circular in nature with a high degree of positive feedback. The accentuated emotional loading on the cortical processes will in time produce a schizoid behavior pattern.

Thus, the first stages of the pathology would be characterized by hyper-alert attitudes and complex perceptual organizations which would later be replaced by increasing loss of contact with reality and finally by random pathological behavior of a generalized hebephrenic nature. The presence of schizophrenic behavioral components in the paranoid individual may be viewed as a consequence of genetic accident plus the personal and private emotional problems acting together in a reactive rather than in a process fashion to produce such behavior.

It is likely that a long range study of paranoid schizophrenics would show that in the early stages of development they would show hypernormal ability to organize perceptual processes. In spite of behavioral pathology, they would not show "schizophrenic" deficit in this area. During the later course of the pathology, a decline in organizational ability would be accompanied by an increase in schizoid symptoms until finally the two pathologies could not be separated on a perceptual basis alone.

Summary

Earlier evidence has indicated that schizophrenics may show impaired ability to perceive visual apparent movement. This paper is a report of an experiment designed to explore the degree of impairment in a large sample of schizophrenics.

There are statistically significant differences between the undifferentiated schizophrenic group and the control group but not between the control group and the paranoid schizophrenic group. These findings support the theory of schizophrenic deficit outlined earlier by Saucer and Deabler (5) but require a new interpretation of the processes of paranoid schizophrenia. A tentative hypothesis concerning the nature of paranoid processes is discussed.

It is concluded that apparent visual movement thresholds may be a useful diagnostic tool but may be better suited to research than to routine diagnostic testing.

Received August 9, 1957.

References

1. Bender, M. B., & Teuber, H. L. Disturbances in visual perception following brain lesions. *J. Psychol.*, 1949, **28**, 223–233.
2. Gallese, A. J. Spiral afterimage as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254.
3. Price, A. C., & Deabler, H. L. Diagnosis of organicity by means of the spiral afterimage. *J. consult. Psychol.*, 1956, **19**, 299.
4. Saucer, R. T. Processes of motion perception. *Science*, 1954, **20**, 806–807.
5. Saucer, R. T., & Deabler, H. L. The perception of apparent motion by organics and schizophrenics. *J. consult. Psychol.*, 1956, **20**, 385–389.
6. Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
7. Werner, H., & Thuma, B. D. A deficiency in the perception of apparent motion in children with brain damage. *Amer. J. Psychol.*, 1942, **55**, 58–67.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Success and Failure in Client-Centered Therapy As a Function of Client Personality Variables[1]

William L. Kirtner and Desmond S. Cartwright
University of Chicago

Cartwright (5) has presented evidence for the existence of a "failure-zone," ranging between 13 and 21 interviews, in client-centered therapy. In a subsequent paper, Taylor (25) presented closely comparable evidence with respect to psychoanalytically oriented psychotherapy.

Cartwright demonstrated that therapeutic outcomes were divided by the failure-zone and suggested that there may be two forms of therapeutic process, identifying these processes as "short" (1–12 interviews) and "long" (13–77 or more interviews). He further hypothesized (5, p. 362) "that certain individual differences between clients give rise to different kinds of therapeutic process."

On the basis of these results and hypotheses, five client groups were distinguished: short success, 1–12 interviews; short failure, 1–12 interviews; failure-zone, 13–21 interviews; long success, more than 21 interviews; and long failure, more than 21 interviews. Utilizing this schema, Kirtner (12) hypothesized that therapy length-by-outcome is related to the personality structures of clients at the beginning of therapy. The present paper reports an investigation of this hypothesis.

Method and Procedure

Subjects

From the research files of the Counseling Center, University of Chicago, all clients were selected for whom the following four conditions held: (*a*) the therapists' rating of outcome on a 9-point success rating scale (22) fell either between 7 and 9 inclusive ("success") or between 1 and 4 inclusive ("failure"); (*b*) the exact number of interviews was recorded; (*c*) a pretherapy Thematic Apperception Test (TAT) protocol was available; and (*d*) a recorded or transcribed first therapy interview was available. Twenty-four clients fulfilled all conditions. Two clients who fulfilled conditions (*a*), (*b*), and (*c*), but not (*d*), were also included, yielding $n = 26$ for this study. All clients were seen by experienced therapists (25 male and 1 female) between 1949 and 1954. The sample included 12 males, 14 females; 10 students, 16 nonstudents; the mean age was 27.5 years and the range 19 to 41 years.

Instruments

Rating scales were developed to measure twelve personality variables. These variables were selected to comprehend: (*a*) some of those variables widely referred to as designating level or pattern of personality integration, e.g., sex role and identification, motivation, impulse life; (*b*) some of those variables particularly relevant to the phenomenological basis of client-centered therapy, e.g., decision-making, rule usage, sense of self as a causative agent (16; 17, pp. 4–5); and (*c*) some of those variables especially associated with previous research in client-centered therapy, e.g., sensitivity to others (23, 24, 26), intrapunitive feelings (9). Upon analysis of the results derived from the application of the 12 scales, 6 scales were eliminated due to functional overlap with the remainder. Numerical ratings made by an independent judge on the 12 scales were intercorrelated. The resulting

[1] This investigation is based upon a dissertation submitted by the senior author to the University of Chicago in partial fulfillment of the requirements for the degree of master of arts. It was supported in part by research grants from the National Institute of Mental Health, National Institutes of Health, U. S. Public Health Service, and from the Ford Foundation.

matrix was inspected for clusters from each of which one scale was selected, yielding four relatively independent scales, numbered I, IV, X, and XI. Scale VIII, while substantially correlated with Scale I, was uncorrelated with the other three selected scales and is therefore reported upon. Scale XII, highly correlated with Scale XI on numerical ratings, is also reported upon because of its distinctive qualitative differentiation between clients. Over the six scales, the range of intercorrelations was $-.24$ to .77. The median correlation was .29.

Following are brief descriptions of the six scales reported in this study. In complete form, all are 5-point, fully anchored scales.[2] A rating of 1 indicates minimal disturbance (optimal adjustment), and a rating of 5 indicates maximal disturbance.

Scale I—Impulse Life

From free flow and expression of generative impulses and ideas, to extreme:

a. fright about impulses but with expression;
b. control or repression;
c. distortion of impulses.

Scale IV—Sense of Capability

From a sense of felt capacity to cope adequately with situations, to a sense of helplessness and lack of internal resources to cope with situations.

Scale VIII—Sex-role and Identification

From direct and open expression of sexuality with good control of feelings, to extreme conflict, uncertainty, and confusion about sex.

Scale X—People—Object Orientation

From the ability to relate deeply both to people and in activities, to:

a. a desperate need for people;
b. avoidance of relating to people with concomitant desperate need for activities;
c. immobilization.

Scale XI—Sense of Comfort and Satisfaction

From a sense of self-comfort and satisfaction with the ability to locate problem sources both internally and externally in good balance, to:

a. intensely driven to search himself for causes of discomfort and dissatisfaction;
b. intensely driven to search outward for causes of discomfort and dissatisfaction;
c. intense hopelessness to locate causes.

[2] Copies of the complete scales and of the raw data from both judges may be obtained without charge from William L. Kirtner, Counseling Center, University of Chicago, 5737 Drexel Avenue, Chicago 37, Illinois, or for a fee from the American Documentation Institute. Order Document No. 5585, remitting $1.75 for microfilm or $2.50 for photocopies.

Scale XII—Punitive Feelings

From a self-appraising person who is not self-reproachful, to:

a. extremely self-punishing person;
b. extreme anticipation of punishment from external source;
c. vacillation between "*a*" and "*b*."

It will be noted that some of the scales include alternate descriptive anchors, indicated by letters *a*, *b*, *c*. Such letter ratings occur only in connection with numerical ratings of 3 or more. As the level of disturbance increases in a given area (as the numerical rating is higher on these scales), the behavior displayed may take on one of several modal characteristics. Salient nonreactive, explosive, or other features of behavior become usual and representatively distinguishing attributes of the person rated. Such salient characteristics or modal patterns of behavior must be discretely described in order to obtain a fully anchored scale of levels of disturbance. As the level of disturbance decreases (as the numerical rating is lower on these scales), the behaviors displayed do not require multiple discrete descriptions; salient nonreactive, explosive, or other features of behavior are neither usual nor representatively distinguishing attributes of the person rated.

While the foregoing paragraph presents the descriptive (i.e., scale-making) problem, the existence of the problem invites explanation. The authors hold the working hypothesis that the individual adopts certain specifiable modes or techniques of behavior in order to control his internal disturbance or anxiety with respect to a given segment of his experience. As anxiety or disturbance intensifies, the techniques required for control of the disturbance, or for maintenance of personality organization, become more drastic and more representatively typical of the individual.

Ratings

Ratings of each client, on each of the six scales, were made on the basis of the pretherapy TAT protocol in conjunction with the recorded or transcribed first therapy interview. In two cases, the recorded first interview was not available, and the TAT protocol alone was used.

Ratings were made independently by the

senior author and by another judge. The latter ("Judge II") knew neither the length nor the success rating of any case.

Reliability

Reliability of the therapist rating of outcome on the 9-point scale has been previously reported (5, 22) as slightly better than $r = .80$. To determine reliability for the numerical ratings on the six scales, all ratings by the two independent judges were correlated using Kendall's tau for ties in both rankings (10). The computational procedure employed was that developed by Cartwright (3). Corrections for continuity were made using Schaeffer and Levitt's (21) modification of Whitfield's correction (27). The routine for computing corrections for continuity and p values was that described by Cartwright (4). Table 1 presents these reliability data, which were considered adequate throughout.

Letter ratings occurred on four out of the six scales. Since such ratings are associated only with numerical ratings of 3 or more on these scales, a long and tedious problem in combinatorics would have to be solved to obtain an exact estimate of reliability, which did not seem justifiable. A fortiori, it did not seem justifiable to work on a solution for the reliability of joint numerical-plus-letter ratings. Instead, simple percentages of agreement were computed for all cases where both judges employed a letter rating, and regardless of numerical rating. Table 2 presents these data and includes, for the reader's information, the number of cases where neither judge used a letter rating.

No statistical analysis of results from letter ratings was made. However, the authors judge the data of Table 2 to warrant descriptive characterization of modal behaviors displayed by subjects rated on the four scales referred to.

Table 1

Reliability Coefficients (Tau) for Numerical Ratings on Six Scales

Scale	Tau	p
I. Impulse life	.65	.01
IV. Sense of capability	.62	.01
VIII. Sex-role and identification	.53	.01
X. People-object orientation	.73	.01
XI. Sense of comfort and satisfaction	.85	.01
XII. Punitive feelings	.83	.01

Table 2

Percentages of Agreement for Letter Ratings on Four Scales

Scale	Number of cases for which letter ratings were used by: Neither judge	Both judges	Agreements: Number	%
I.	3	22	20	91
X.	0	26	23	88
XI.	0	26	25	96
XII.	1	25	22	88

Analysis

The sample of 26 cases was divided into length-by-outcome groups according to the schema given. Table 3 summarizes descriptions of the resulting groups.

Regrettably for the research, only one long-failure case was among those available. This group was therefore eliminated from subsequent analysis. The remaining groups were examined in all possible pairs on each of the six scales as numerically rated by Judge II. All comparisons were made using tau for a dichotomy in one ranking as described by Kendall (10, pp. 32–34). The significance of differences between groups was computed in the manner previously indicated for the reliability analysis. In order to show the direction of differences between pairs of groups, mean values for each group on each scale were also computed.

Results

Numerical Ratings

Table 4 presents means for groups on each of the six scales and the p values of the differences between groups.[8]

Rejecting the null hypothesis if $p < .05$, each success group shows a lesser degree of

[8] See Footnote 2.

Table 3

Summary Description of Groups

Group	Number of interviews	Range of outcome ratings	*N*	Male	Female	Student	Non-student
Short failures	<13	1–4	10	4	6	4	6
Short success	<13	7–9	4	0	4	3	1
Failure-zone	13–21	1–4	7	5	2	1	6
Long failure	>21	1–4	1	1	0	1	0
Long success	>21	7–9	4	2	2	2	2

disturbance than either failure group on Scale I. Also, the short-success group shows a lesser degree of disturbance than either failure group on Scale VIII.

Of special interest are results setting a given group apart from the other three groups. We reject sets of three null hypotheses if in each instance $p < .10$. On this basis, Table 4 contains the following results: the short-success group shows less disturbance than any other group on Scales I and VIII; the short-failure group shows more disturbance than any other group on Scale IV; and the failure-zone group shows more disturbance than any other group on Scale VIII.

Letter Ratings

Table 5 presents the number of letter ratings agreed upon by both judges for four scales with respect to the four groups.

The most interesting feature of Table 5 is that success groups appear to differ from failure groups on Scales X, XI, and XII. While there were no differences on these scales as numerically rated for degree of disturbance, differences in typical mode of controlling inner disturbance were observed by both judges.

Distinguishing Descriptions

The results and observations mentioned in connection with Tables 4 and 5 are given embodiment in descriptions distinguishing the groups. Descriptions employ the language of those scales, and only those scales, for which either statistically significant differences or clear qualitative observations have been made.

The short-success group shows a generally higher level of personality integration than that descriptive of other groups. In particular, they are rather open to their impulse life (I), and, while showing some distortion and uncertainty, they are a good deal less confused about their sex role than other groups (VIII).

Though similar to the short-success group in many respects, the long-success group shows a somewhat lower level of personality integration in general. In particular, they are rather anxious about their impulse life (I)

Table 4

Significance Levels of Differences Between Groups on Six Scales, and Means Indicating Direction of Differences

Scale	Means of Groups[b]				*p* of Differences[a]					
	SS	LS	SF	FZ	SS vs. LS	SS vs. SF	SS vs. FZ	LS vs. SF	LS vs. FZ	SF vs. FZ
I.	2.00	3.00	4.10	4.43	<.10	<.01	<.01	<.01	<.01	>.20
IV.	3.00	3.25	4.30	3.71	>.20	<.02	>.20	<.03	>.20	<.07
VIII.	2.50	3.75	3.80	4.57	<.07	<.02	<.02	>.20	<.10	<.02
X.	3.50	3.75	4.10	4.00	>.20	>.20	>.20	>.20	>.20	>.20
XI.	3.75	4.50	3.90	3.57	<.15	>.20	>.20	>.20	<.08	>.20
XII.	3.25	4.25	3.80	3.85	<.19	>.20	>.20	>.20	>.20	>.20

[a] Two-tailed tests throughout.
[b] SS—short success; LS—long success; SF—short failure; FZ—failure-zone.

Table 5

Letter Ratings Agreed on by Both Judges

	Short success			Long success			Short failure			Failure zone		
Scale	*a*	*b*	*c*	*a*	*b*	*c*	*a*	*b*	*c*	*a*	*b*	*c*
I.				4			2	3	4		5	1
X.	4			4			1	4	4	2	1	2
XI.	3			4				8	2		6	1
XII.	3			4				8	1		4	1

and rather confused about the meaning and place of sex in their identity (VIII).

Both success groups differ from each failure group in showing less disturbance in their impulse life (I). Letter ratings show the success groups typically displaying a strong need for people to relate to in order to feel worthwhile (X), seeking internally for the cause and resolution of felt discomfort (XI), and rather strongly condemning, berating, and disvaluing themselves (XII).

The short-failure group shows a generally low level of personality integration. Especially distinguishing is their extreme underlying sense of incapacity to deal with their life situations (IV).

The failure-zone group also shows a generally low level of personality integration. Especially distinguishing is their extreme conflictedness about and rejection of their sex role (VIII).

Both failure groups show greater disturbance than either success group on several scales. They are extremely disturbed in their impulse life (I), with expressions frequently occurring in over-controlled or distorted modalities, as indicated by letter ratings. Overcontrol seems somewhat more typical for the failure-zone group. Letter ratings show the failure groups typically looking outward for the cause and resolution of their discomfort (XI), and typically anticipating punishment from external sources (XII).

Discussion

The foregoing results seem unambiguously consistent with the general hypothesis motivating the study: that therapy length-by-outcome is related to the personality structure of clients at the beginning of therapy. While caution must be invoked due to the small sizes of the groups studied, the results add to the growing body of knowledge about factors which favor or limit therapeutic change in client-centered therapy. Earlier work in this area has been summarized by Rogers (**17**, pp. 423–424, 427–428).

It may be noted that the present approach constitutes a synthesis of two trends in previous work on the general problem. One trend has dealt with the relations between pretherapy characteristics of clients and eventual ratings of improvement or outcome (e.g., **1**, **2**, **8**, **11**, **15**, **19**, **26**). The other trend has dealt with the relations between pretherapy characteristics of clients and their length of stay in treatment (e.g., **7**, **13**, **14**, **18**, **20**). To our knowledge, only one previous investigator, Dana (**6**), has attempted to predict improvement and length of treatment jointly. Due to the findings of Cartwright (**5**) and Taylor (**25**), a fusion of the two main trends into systematic emphasis upon length-by-outcome groupings now seems necessary. More generally, it is necessary to predict both amount and rate of change constituting improvement. While something of a beginning has been made in this direction by Dana's results and by those of the present study, much more needs to be done.

Summary

Twenty-six clients, who had completed a course of client-centered therapy at the Counseling Center, University of Chicago, were divided into five groups on the basis of length-by-outcome:

Short success (1–12 interviews)—*N* equals 4
Short failure (1–12 interviews)—*N* equals 10
Failure-zone (13–21 interviews)—*N* equals 7
Long success (more than 21 interviews)—*N* equals 4
Long failure (more than 21 interviews)—*N* equals 1

Twelve personality variables were selected, and 5-point rating scales were constructed to describe both intensity and mode of behavior with respect to each of the 12 variables. These scales were applied independently by two judges to the pretherapy Thematic Apperception Test protocol in conjunction with

the recorded or transcribed first therapy interview in each case. Interjudge agreement was deemed adequate throughout.

The ratings made by an independent judge were intercorrelated, and the resulting matrix was inspected for clusters from each of which one scale was selected, yielding four relatively independent scales. Two further scales were reported upon because, though substantially correlated with one of the four selected scales, they are uncorrelated with the others and point up interesting and distinctive differences between client personality structure.

It was thus found that therapy length-by-outcome is related to the personality structure of clients at the beginning of therapy. The most marked differences found on these scales were those between success groups and the failure groups.

Received July 31, 1957.

References

1. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.
2. Barron, F. Some test correlates of response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 235–241.
3. Cartwright, D. S. A computational procedure for tau correlation. *Psychometrika*, 1957, **22**, 97–104.
4. Cartwright, D. S. A note concerning Kendall's tau. *Psychol. Bull.*, 1957, **54**, 423–425.
5. Cartwright, D. S. Success in psychotherapy as a function of certain actuarial variables. *J. consult. Psychol.*, 1955, **19**, 357–363.
6. Dana, R. H. The effects of attitudes towards authority on psychotherapy. *J. clin. Psychol.*, 1954, **10**, 350–353.
7. Gallagher, J. J. A comparison of clients who continue with clients who discontinue client-centered therapy. In Psychotherapy Research Group (Ed.), *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State College, 1953. Pp. 21–38.
8. Gallagher, J. J. Test indicators for therapy prognosis. *J. consult. Psychol.*, 1954, **18**, 409–413.
9. Haimowitz, Natalie R., & Haimowitz, M. L. Personality changes in client-centered therapy. In Werner Wolff (Ed.), *Success in psychotherapy*. New York: Grune & Stratton, 1952. Pp. 63–94.
10. Kendall, M. G. *Rank correlation methods*. London: Griffin, 1948.
11. Kirkner, F. J., Wisham, W. W., & Giedt, F. H. A report on the validity of the Rorschach Prognostic Rating Scale. *J. proj. Tech.*, 1953, **17**, 465–470.
12. Kirtner, W. L. Success and failure in client-centered therapy as a function of personality variables. Unpublished master's thesis, Univer. of Chicago, 1955.
13. Kotkov, B., & Meadow, A. Rorschach criteria for predicting continuation in individual psychotherapy. *J. consult. Psychol.*, 1953, **17**, 16–20.
14. Lutes, B. C. A factorial study of an attrition group in client-centered therapy. Unpublished master's thesis, Univer. of Chicago, 1952.
15. Mindess, H. Predicting patients' responses to psychotherapy: A preliminary study designed to investigate the validity of the Rorschach Prognostic Rating Scale. *J. proj. Tech.*, 1953, **17**, 327–334.
16. Rogers, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
17. Rogers, C. R., & Dymond, Rosalind F. (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
18. Rogers, L. S., Krauss, Joanne, & Hammond, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.*, 1951, **15**, 368–371.
19. Rosenberg, S. The relationship of certain personality factors to prognosis in psychotherapy. *J. clin. Psychol.*, 1954, **10**, 341–345.
20. Rubinstein, E. A., & Lorr, M. A comparison of terminators and remainers in outpatient psychotherapy. *J. clin. Psychol.*, 1956, **12**, 345–349.
21. Schaeffer, M. S., & Levitt, E. E. Concerning Kendall's tau, a nonparametric correlation coefficient. *Psychol. Bull.*, 1956, **53**, 338–346.
22. Seeman, J. Counselor judgments of therapeutic process and outcome. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954. Pp. 99–108.
23. Sheerer, E. T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169–175.
24. Stock, D. An investigation into the interrelations between the self-concept and feelings directed toward other persons and groups. *J. consult. Psychol.*, 1949, **13**, 176–180.
25. Taylor, J. W. Relationship of success and length in psychotherapy. *J consult. Psychol.*, 1956, **20**, 332.
26. Tougas, R. R. Ethnocentrism as a limiting factor in verbal therapy. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954. Pp. 196–214.
27. Whitfield, J. W. Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika*, 1947, **34**, 292–296.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Performance of Psychoanalytic and Client-Centered Therapists in an Initial Interview[1]

Hans H. Strupp[2]
School of Medicine, University of North Carolina

As a technique for producing changes in the personality structure of a patient, psychotherapy does not exist apart from the person of the therapist. While it is axiomatic that the therapeutic instrument comprises the therapist, his personality, theory, and techniques, little systematic attention has been paid by researchers to these variables.

In a sense, the therapist is a reader of imaginary dials who acts "therapeutically" upon the information indicated by these measuring instruments within himself. When supplied with certain information, say, the behavior of a patient, he will register this information in certain ways, and he will react to it in certain other ways. The therapist's theory, experience, and personality determine the kinds of dials he uses as well as their calibration. His therapeutic actions are their reflection.

In order to obtain empirical data on the calibration of the therapist as a clinical instrument, we can supply a certain identical input (patient behavior) and note (*a*) how this input is perceived and evaluated; (*b*) what deductions and inferences are drawn from it; and (*c*) how these processes are translated into therapeutic action. On the hypothesis that this testing procedure will disclose differences among therapists, one can investigate systematic influences which might produce the discrepancies. For example, therapists may use different sets of measuring instruments whose readings have no relation to each other. They may use similar instruments, but the calibration may differ. They may interpret differentially the meaning of the same pointer reading. Any one of these possibilities, plus many others, may result in divergent therapeutic communications.

This analogy is of course a gross oversimplification of the therapist's activity in the therapeutic situation, but it may help to make explicit the objectives of the larger investigation of which this paper forms a part. The purpose was to study the therapist's performance in a first interview and to ask: If a group of therapists observe the same patient, how do they perceive the clinical problem? What kinds of evaluations result from their perceptions? How are their perceptions and evaluations related to their communications? And, can differences in their performance be accounted for in terms of common underlying variables? Specifically, this paper is concerned with a comparison between psychotherapists (psychologists) following psychoanalytic principles and therapists following the client-centered theory of Rogers.

Method and Procedure

To study the performance of psychotherapists under comparable conditions, a sound film of a first interview was used as an experimental model. The film was produced by the Veterans Administration for training purposes and was entitled "A Clinical Picture of

[1] This paper was presented at the 1957 meeting of the Eastern Psychological Association.

This research, part of a larger project supported by a research grant (M-965) from the National Institute of Mental Health, Public Health Service, was conducted while the author was affiliated with The George Washington University, School of Medicine.

[2] The author is indebted to Rebecca E. Rieger for her assistance and valuable comments.

Claustrophobia." [3] It was described as an unrehearsed interview from which no significant portions had been deleted. The motion picture was adapted for experimental purposes by introducing interruptions at 28 preselected points identified by a uniform title "What would you do?" followed by a number from 1 to 28.

Synopsis. The patient, a middle-aged man, complains of intense anxiety and gastrointestinal symptoms which appear to be psychogenic. He relates an episode of a panic reaction on a crowded street car. Following the incident he found himself unable to go out of the house unless he was accompanied by his wife, an invalid of many years. As the interview progresses, he dwells on an intense conflict with his mother, a strong-willed woman, and speaks at length about his "bad luck" in all phases of his life. The father had died when the patient was young, leaving him with family responsibilities for which he was unprepared. He feels that "I had three strikes against me before I came to the plate." He speaks freely, and his anger and resentment against his mother, brother, and people in general build up. The patient appears hostile, demanding, and dependent; he verbalizes a feeling that the world owes him a living. While it is not clear that he had psychotherapy before, he makes mention of previous interviews. He speaks of having obtained some "insight," of which his rationalizations and projections give little evidence. The interview ends somewhat abruptly after he has made a passionate plea for help, couched in a demand for reassurance.

The interviewer was an inexperienced resident in psychiatry who essentially permitted the patient to tell his story. He frequently repeated a word or phrase of the patient's narrative but focused very little. At times, he asked informational questions which seemed to have the effect of diverting the patient from an emotionally charged topic. The interviewer's attitude may be described as benevolent but rather distant. He was passive, offered no reassurance and seemed ill at ease. No clues were given about the outcome of the interview.

The film was shown to groups of therapists who were asked to assume that they were interviewing the patient, and to record in writing their communications to the patient while the film sequence was interrupted. Each pause lasted 30 seconds, which provided most therapists ample but not excessive time to make their comments. Respondents were free to respond or to remain "silent." The experimental version of the film ran continuously and lasted approximately 50 minutes.

Following the film showing, therapists were requested to complete a comprehensive questionnaire on diagnostic impressions, treatment plans and goals, formulations of the patient's emotional dynamics, problems in treatment, estimates of the degree of anxiety, emotional maturity, social adjustment, prognosis, the respondent's attitude toward the patient, etc. In addition, there was a biographical information blank containing 24 questions about the therapist's training, theoretical orientation, and current therapeutic activities.

The experimental procedure usually required about two hours.

The Samples

Cooperation was obtained from psychiatric and psychological societies, hospitals, training centers, and therapeutic institutes. Data were collected in Washington, Baltimore, New York, Perry Point, and Chicago. The total sample comprised 235 psychotherapists, 134 of whom were members of the medical profession (analysts, psychiatrists, and residents in psychiatry); 17 were psychiatric social workers. This paper is concerned with the responses of 14 psychologists who described their theoretical orientation as unqualifiedly Rogerian (8 were affiliated with the Counseling Center of the University of Chicago). This group is compared to 64 psychologists who professed to follow psychoanalytic principles broadly defined. Actually, they selected a total of 93 "labels," which were distributed over 14 categories; 72% of the choices were accounted for by: psychoanalytic orientation (general or orthodox), Sullivan, and eclectic.[4]

Summary statistics on age, experience, and personal analysis are presented in Table 1. The following data shed some further light on the composition of the samples:

[3] The "Psychotherapeutic Interviewing Series" was directed by Jacob Finesinger and Florence Powdermaker. The film under discussion, Part V of the series, was filmed in 1952. I wish to express my appreciation to the Psychiatry and Neurology Service of the Veterans Administration for granting permission to use the film for experimental purposes.

[4] Grateful acknowledgment is made to all psychologists who gave generously of their time to make this study possible. Thanks are also due the following organizations: VA Benefits Office, Washington, D. C.; VA Hospital, Perry Point, Md.; Maryland Psychological Association; Postgraduate Center for Psychotherapy, New York, N. Y.; Counseling Center, University of Chicago.

Table 1

Summary Statistics of Age, Experience, Personal Analysis

Statistic	Sample A[a]			Sample R[b]		
	N[c]	Mean	*SD*	*N*[c]	Mean	*SD*
Age	65	35.6 yrs.	8.7	13	33.4 yrs.	6.4
Experience in therapy	62	3.3 yrs.	3.1	12	3.0 yrs.	2.4
Personal analysis	33	291 hrs.	198	9	142 hrs.	101

[a] Psychoanalytically oriented psychologists.
[b] Rogerian psychologists.
[c] Fluctuations reflect absence of complete information and/or not having had personal analysis.

Two-thirds of the psychoanalytically oriented psychologists (Sample A) held the Ph.D. degree, as compared with 36% of the psychologists who followed Rogers' theory (Sample R).

Sixty per cent of the members in both groups said they had a fair amount or a great deal of experience in therapy and considered themselves competent therapists. All members of Sample R said they practiced client-centered therapy exclusively. Sample A showed a wider distribution, but most replies fell within the broad category of psychoanalytically oriented psychotherapy. Most Rogerians indicated that they had received very little training in psychoanalytic principles and techniques; 69% of Sample A claimed "a fair amount" or "a great deal."

Sixty-four per cent of Sample R had received their personal analysis within the client-centered framework. The responses of members of Sample A who indicated that they either had completed or were in the process of undergoing personal analysis (more than half the group) suggested that their analyses were based upon psychoanalytic theory.

Approximately one-half of the members in both groups stated that they devoted more than 25% of their time to the practice of psychotherapy. Almost two-thirds of Sample R specified their therapy as being primarily with neurotic patients. Thirty-eight per cent of Sample A responded similarly, but another third said that their patients represented a wider range of personality disorders.

Sample A appeared to treat their "average patient" for longer periods than Sample R: 38% of the former group described the length of therapy as exceeding one year, as contrasted with 14% in the latter. On the other hand, Rogerians predominantly saw the "average patient" twice a week, whereas the modal frequency for Sample A was closer to one weekly session.

Results

Therapists' responses to the film, that is, communications to the patient, were quantified by a multidimensional system of analysis which has been described previously (11). Questionnaire responses to each item were coded and the frequencies tabulated. In view of the small size of Sample R, response categories within an item had to be combined. Typically, this resulted in 2×2 tables which were tested by means of chi square (corrected for continuity).[5]

Clinical Evaluations and Therapist Attitudes

Responses to the question of diagnosis revealed no significant differences. It seems noteworthy, however, that no member of Sample R chose such labels as psychopath, schizoid, paranoid, or character disorder, which were used by appreciable minorities of Sample A, as well as by psychiatrists whose response distributions have been presented elsewhere (12). Asked to state the patient's major defense mechanisms, Sample R named a relatively smaller number than Sample A (ratio of 1:3), but the distributions did not appear to differ significantly. Nor did the samples differ in terms of qualitative descriptions of the patient's emotional dynamics.

Evaluations of the degree of ego strength, anxiety, insight, social adjustment, and disturbance disclosed no statistically significant differences.

The members of Sample R judged the prognosis with therapy to be more favorable than the members of Sample A ($p = < .05$): no

[5] A table showing the response distributions of the two samples in percentages and giving probability values for $p = < .10$ has been deposited with the American Documentation Institute. Order Document No. 5587, remitting $1.25 for microfilm or $1.25 for photocopies.

Rogerian therapist considered it unfavorable. By contrast, an unfavorable prognosis was given by 40% of the therapists in Sample A, who were in closer agreement with the predictions made by psychiatrists.

A larger proportion of Rogerian therapists professed a positive attitude toward the patient than analytically oriented psychologists ($p = < .02$). Members of the University of Chicago Counseling Center unanimously described their attitude as positive. Those responding negatively often stated that their attitude was influenced by the patient's hostility, anger, and dependency; those responding positively stressed his personality assets or saw him as "a patient with a problem," a response often given with some empathy.

Plans, Goals, and Difficulties in Therapy

Sample R showed greater reluctance to set up therapeutic goals than Sample A ($p = < .001$), but the two samples did not differ significantly as to estimated length of treatment or time needed to accomplish symptom relief. It seems pertinent to note that psychoanalysts were less optimistic than either of the psychologist groups.

Concerning difficulties expected in treatment, Sample R therapists were significantly less likely to mention: the patient's parasitic attitudes, poor motivation for therapy, acting out, or the possibility of psychosis ($p = < .02$). Nor were countertransference or problems of technique mentioned with any degree of frequency by Sample R ($p = < .01$). One-half of Sample R anticipated countertransference problems with this patient, as compared with almost 80% in Sample A; twice as many Rogerians answered the question in the negative ($p = < .07$). No Rogerian mentioned areas in the patient's living which should be avoided, at least early in therapy, whereas numerous areas were mentioned by Sample A therapists ($p = < .001$). Similarly, Rogerians were less likely to name areas in the patient's living which should be singled out for focus in therapy ($p = < .001$).

Rogerians either declined to specify attitudes or behaviors which the therapist should encourage in therapy with his patient or stressed the expression of feelings. Sample A therapists were more likely to stress: a sense of responsibility, increased socialization, relating feelings and symptoms to interpersonal situations, etc. ($p = < .02$).

Conversely, Sample A therapists were more likely to discourage attitudes and behaviors, such as: intellectualization, obsessive ruminations, self-pity, self-depreciation, helplessness, refusal to accept responsibility, demanding attitudes, and acting out; Rogerians, by comparison, tended to say they would discourage nothing or leave it to the patient ($p = < .001$).

No member of the Rogerian group advocated strictness by the therapist, as contrasted with more than a third of the members of Sample A who considered strictness therapeutically desirable ($p = .02$). A question about therapist's activity or passivity did not differentiate the two groups, possibly because the terms of reference were not clearly defined. Free association was discouraged by an appreciable proportion of Sample A therapists, but not by Sample R ($p = < .06$).

A question was asked about clues in the patient's verbal or nonverbal behavior which might not be used immediately but kept in mind for possible later reference. All members of Sample A who answered the question (85%) mentioned one or more clues. Four Rogerians specifically stated "none" ($p = < .001$). Rogerians commented about the patient's feelings and attitudes but, in contrast to Sample A, paid less attention to such clues as: gestures, bodily movements, manner of speaking, the patient's past or present interpersonal relations with his mother, wife, brother, or father—with or without inference about their dynamic significance.

Descriptions of the course of the transference did not diverge significantly, but members of the two samples disagreed about approaches to the problem ($p = < .001$). Rogerians predominantly encouraged or fostered a "corrective emotional experience" by conveying respect or clarifying feelings. Members of Sample A, on the other hand, tended to advocate interpretation and other procedures, such as reassurance and firmness, in which the therapist uses his role to induce changes in attitudes and actions.

Responses to a question dealing with "acting out," an aspect of the transference prob-

lem, revealed a similar pattern: no statistically significant differences with respect to its occurrence, but divergence in terms of "handling." Here, too, Rogerians recommended understanding, clarification, and reflections; analytically oriented therapists preferred an interpretive approach ($p = < .001$).

Evaluation of and Attitude Toward the Film Therapist

A number of questions were included to study similarities or differences between the film therapist's and the respondent's approaches to interviewing. It was hoped that in this way it might be possible to obtain a clearer picture of the respondent's own procedure.

Sample R therapists were more definite in their assertion that they would have conducted the interview "in a very different manner" ($p = < .02$). They dissociated themselves from his approach and tended to evaluate his performance as "a very inadequate job"; the majority of therapists in Sample A considered his performance reasonably adequate ($p = < .01$).

Within each sample, there was notable disagreement as to whether the therapist had been too passive or given insufficient support, but the intersample comparisons were nonsignificant.

Rogerians and psychoanalytically oriented therapists disagreed very markedly about the amount of time they would have devoted to obtaining data on the patient's life history: 75% of Sample R stated that they would have spent somewhat less or considerably less time than the film therapist; almost one-half of Sample A said they would have devoted somewhat more or considerably more time ($p = < .001$). The transcript shows that about 10% of the therapist's interventions can be construed as case history questions in the usual sense. These questions were usually asked in the context of the patient's story, and attempted to clarify a time, place, or interpersonal event. On the other hand, the therapist frequently repeated a word or phrase in a questioning tone of voice, apparently to obtain further data on a feeling or attitude.

The two samples revealed a statistically highly significant difference in terms of their attitude toward the therapist. The majority of Rogerian therapists described their attitude as negative, whereas most analytically oriented therapists professed a positive or neutral attitude ($p = < .001$).

A supplemental question disclosed that the respondents' criticisms were directed at the film therapist's technique as well as at his person. Prominently mentioned by both samples were the following objections: Therapist was not in control of the situation, engaged in too much activity but not of the "right" kind or not at the "right" time, too much interference with patient's talk, too much "accenting" (repeating last word), wrong emphasis, no plan. Rogerians criticized him especially for failing to focus on the patient's feelings and attitudes, and for a lack of empathy.

Interviewing Techniques

Analyses of therapists' communications to the patient disclosed that Rogerians strongly favored the reflection-of-feeling technique, which comprised 67% of their responses. The comparable percentage for analytically oriented psychologists was 17. The technique preferred by Sample A consisted of exploratory questions (40%). The number of "silent" responses given by Sample R was significantly smaller than that of Sample A. Communications showing a high degree of inference or initiative were rare in Sample R; Sample A had 3 to 8 times as many responses in these ranges.

As might be expected, Sample R therapists adhered quite consistently to the reflection-of-feeling technique, even when the patient pleaded for a direct answer from the therapist. Many analytically oriented therapists responded to this plea with a direct answer.

Discussion

In interpreting these findings it must be kept in mind that the sample of Rogerian therapists was small. This deficiency was partly mitigated, however, by the fact that this group was compared with a considerably larger sample of analytically oriented therapists to whom they were roughly comparable in terms of training and level of experience. Furthermore, on a number of questionnaire

items, Sample R revealed almost complete unanimity which undoubtedly reflected homogeneous training in a single setting (The University of Chicago Counseling Center); Sample A, on the other hand, was made up of therapists from a variety of backgrounds. It may be noted that psychoanalysts (members of the American Psychoanalytic Association), whose training might be considered reasonably homogeneous, varied more widely in their responses.

The results showed a number of pronounced differences in approach which highlighted discrepancies between therapists subscribing to psychoanalytic principles and the client-centered framework. Both positions are too well known to require exposition in this context. It seems appropriate, however, to quote what Rogers considers to be the distinctive emphases of client-centered therapy:

> A second distinctive aspect [the first is the emphasis on research] is the deep confidence in the capacity of the individual, which client-centered therapists have found to be justified by their experience. Rather than seeing man as essentially a destructive animal [the reference is evidently to the death instinct postulated by Freud late in his career, one of the most disputed and least accepted tenets in contemporary analytic thinking], with the primary problem of the control of his impulses, we have found man to be essentially constructive, if he can be released from his defensiveness. Our experience is that, to the extent that he can fully accept his experience and himself, he becomes satisfactorily self-controlled. Our view of the basic nature of man differs sharply from those of most therapeutic orientations, which believe that guidance and control must be arranged by the expert.
>
> A third point of difference relates to a difference in theory. The theory of therapy which is developing in client-centered therapy stresses the importance of the immediate moment of experience in the relationship, not the genetic causation of behavior in the client, nor the formulation of some theoretic picture in the mind of the therapist (**9**, p. 207).[6]

[6] It is difficult to see how the terms "guidance and control," in the sense that Rogers seems to use them, do in fairness characterize the mainstream of contemporary psychoanalytic thinking. A semantic confusion seems to arise from the conception that the psychotherapist who functions in the role of a professional expert *ipso facto* manipulates the patient or controls him, and that guidance in the educational sense is synonymous with authoritarianism. Compare, in this connection, the formulations of a prominent psychoanalyst regarding therapeutic goals and the part played by the therapist:

> Treatment, of course, is aimed at the solution of the patient's difficulties in living and the cure of his symptomatology. Ideally these therapeutic goals will be reached by the growth, maturation, and inner independence of the patient. . . . This goal will also be actualized by the development of his capacity for self-realization, his ability to form durable relationships of intimacy with others, and to give and accept mature love (**3**, p. 34).
>
> . . . security and inner independence of the authoritarian values attributed to the conventional requirements of our culture are indispensable for the therapist who wants to guide his patients successfully toward finding out about the degree of cultural adjustment which is adequate to their personal needs (**3**, p. 33).
>
> The psychiatrist's respect for his patients will also help him to safeguard against the previously mentioned mistake of assuming an attitude of personal "irrational authority" instead of listening and conducting therapy in the spirit of collaborative guidance. This irrational authoritarian behavior will be harmful not only because it interferes per se with the patient's tendency toward growth and maturation but also, and more important, because it constitutes a traumatic repetition of the authoritarian aspects of the cultural pattern of behavior in general and of the parental pattern in particular, to which most mental patients have been harmfully subjected in their past (**3**, p. 17).

Rogers' conception of the "ideal" therapist is that of a participant in an interpersonal situation whose primary function is to empathize with and understand the patient within his own frame of reference. In his third point of difference, Rogers asserts that the experiential element in the psychotherapeutic situation is more important than the formulation of some theoretic picture in the mind of the therapist. This statement carries with it the implication that the therapist cannot adequately empathize with the patient, thereby helping to actualize the immediate moment of experience, if he attempts to understand rationally what is going on. Rogers has elaborated on this alleged dilemma in another paper (**8**). I, for one, know of no evidence that the therapist's private thoughts about the nature of the patient's emotional problem, its severity, etiology, etc., in and of themselves preclude or interfere with empathy, nor that the exclusion of the therapist's rational processes assures it. I feel that any psychotherapist does more than empathize with his patient: whether he interprets a feeling or attitude in terms of current or past interpersonal relationships, or whether he selects a particular feeling or attitude for reflection, he seems to follow a design for action dictated by his theory.

Psychotherapy based on psychoanalytic principles does differ from other forms of psychotherapy in its emphasis on giving the patient a rational understanding of his interpersonal processes as they emerge in the therapeutic relationship. The emotional experience remains uppermost, but, in addition, the patient gains a conscious appreciation of his own contribution to his difficulties in interpersonal relations.

his (the patient's) frame of reference. From this vantage point, the therapist is imposing an external frame of reference if he attempts to diagnose, assess, or prognosticate. The same dictum would apply to hypotheses or hunches the therapist may have about the etiology of the patient's conflict or how best to help him through psychotherapy. By this definition, the therapist who explicitly or implicitly goes beyond the patient's phenomenological self-descriptions is not client-centered.

Certain objections voiced by Rogerian therapists regarding the experimental procedure are understandable from this point of view. While they cooperated willingly, they made it clear that ordinarily they do not work this way.

On this basis, too, the following results are consistent with the client-centered viewpoint: a reluctance to make a diagnosis beyond the general label "anxiety," to formulate therapeutic goals, to mention dangers that might be encountered in therapy or cautions that would have to be observed, to encourage or discourage specific attitudes or behaviors in the patient, to treat him with strictness, to discourage the free expression of feelings (free association), or generally to consider problems of technique or "handling." Rogerians were also disinclined to make inferences about the patient's behavior in the interview situation, the etiology of his disturbance, or the dynamic import of his behavior. Finally, they de-emphasized case history questions.

One of the sharpest discrepancies between client-centered and analytically oriented therapists related to the question of prognosis: *no* Rogerian gave an unfavorable estimate, whereas 40% of Sample A took a pessimistic view. How can this finding be explained?

First, it must be considered that our knowledge of neurotic disorders and psychotherapy is not sufficiently developed to permit specific predictions about the outcome of treatment even when a great deal is known about the patient. Such predictions undoubtedly become increasingly accurate as therapy progresses, and the therapist comes to learn more about the nature of the conflict, its ramifications, its history, and the patient's current level of functioning. Thus, a prognostic estimate based on a single interview may be little more than a guess, which may reveal more about the therapist than about the patient.

Second, the key issue in this comparison seems to be the high degree of uniformity among Rogerians which was not paralleled by the analytically oriented therapists whose responses appeared to be more "normally" distributed. If therapists were talking about a specific event, the predictions of some obviously must be right and the predictions of others must be wrong. Unquestionably, therapists were not talking about a specific event but about a therapeutic outcome concerning which a wide range of ideas and opinions were entertained.

In an attempt to shed some light on this result, I propose to examine the following alternatives: (*a*) Rogerians gave a more favorable prognosis because they perceived the patient as less disturbed; (*b*) Rogerians planned for more modest therapeutic goals and therefore gave a more optimistic prognostic estimate; (*c*) prognostic estimates may be relatively unrelated to the presenting problem or the therapist's past successes with similar patients, but reflect more directly the therapist's attitude, his theoretical orientation, or combinations of these.

(*a*) The available evidence indicates that therapists in both samples tended to regard the patient as seriously disturbed. They were also reasonably agreed that he suffered a great deal of anxiety, that he was emotionally immature, socially maladjusted, and that he had relatively little insight. There is no reason to believe, therefore, that Rogerians judged the patient's difficulties as less severe.

(*b*) It is difficult to judge similarities or differences in therapeutic goals because of the vagueness of the terms of reference. Rogerians tended to advocate greater self-acceptance; analytically oriented therapists tended to strive for some insight. However, the latter group generally did not recommend the more intensive forms of psychotherapy usually associated with psychoanalysis. On this basis, it seems fair to say that the Rogerians' more favorable prognostic estimates were not a function of more limited therapeutic goals.

(*c*) It has been reported previously (12) that there was a highly significant statistical relationship between psychiatrists' prognostic estimates and their self-ratings of their conscious attitude toward the patient ($r = .39$). For psychoanalytically oriented psychologists, the correlation was $r = .42$. Among Rogerians, the invariance between attitude and prognosis was almost complete. Eight therapists at the University of Chicago Counseling Center expressed a positive attitude and rated the prognosis favorable.

In the case of psychiatrists, it was found that both attitude and prognosis were significantly associated with other clinical evaluations and treatment plans. In addition, therapists' attitudes tended to be reflected in the emotional tone and the degree of empathy of their communications. These findings suggested the interpretation that clinical assessments tended to reflect, in part at least, the therapist's emotional reaction to the patent.[7] On the other hand, if the therapist had undergone personal analysis, his empathy tended to be greater irrespective of his conscious attitude. In other words, personal analysis seemed to introduce some distance between a negative reaction to the patient and the therapist's ability to empathize with him. It is possible, then, that the therapist's conscious attitude toward the patient is less important than what he does about it.[8]

[7] It seems reasonable to assume that clinical evaluations, treatment plans, and goals are a function of the diagnosis. Comparisons of therapists (based on the total sample) giving a "pure" diagnosis of Hysteria, Obsessive-compulsive, Paranoid, and Anxiety showed a number of statistically significant differences in terms of clinical evaluations, etc. (12). It was pointed out in the same connection that the choice of certain diagnostic labels (paranoid, psychopath) may signify a subtle value judgment which in turn may be determined by the therapist's attitude toward the patient. The data seemed to favor this conclusion.

[8] It should be recalled that the operational definition of therapist's conscious attitude toward the patient, as used in this investigation, was a self-rating on a five-point scale. Apart from considerations of reliability, this rating was undoubtedly influenced by therapists' willingness to express their genuine conscious attitude, unconscious distortions, etc. It is difficult to say to what extent such ratings represent countertransference reactions even if there were greater agreement about the term (5). It seems that Gitelson's (4) description of reactions to the patient as a whole, which he terms transferences of the analyst as distinguished from reactions to partial aspects of the patient (countertransference), fits the present ratings reasonably well:

> It is my impression that total reactions to a patient are *transferences* of the analyst to his patients and are revivals of ancient transference potentials. These may be manifested in the over-all attitude toward patients as a class or may exacerbate in the "whole response" to particular patients. These attitudes may be positively or negatively toned. They are likely to manifest themselves early in the contact with a patient and determine the tendency of the analyst towards the whole case (4, p. 4).

To integrate a workable therapeutic relationship, the therapist, according to psychoanalytic theory, must convey an attitude of acceptance, patience, respect, etc. He must also be able to empathize (1, p. 580). This does not mean that he must have or should have a positive attitude toward *all* aspects of the patient's behavior. Rather, it is recognized that the therapist, like other people, may experience emotional reactions. However, he is enjoined to be vigilant to make sure that they do not interfere with the therapeutic process and if they do, to seek clarification through self-analysis or other means.

The emphasis on diagnosis and etiology, according to analytic theory, determines at least partially the therapist's approach to therapy. Reich says:

> It should be clear that one approaches an aggressive patient unlike a masochistic one, a hyperactive hysteric unlike a depressive one, that one changes one's attitude to one and the same patient according to the situation, that, in brief, one does not behave neurotically oneself, even though one may have to deal with some neurotic difficulties in oneself.
>
> One cannot give up one's own individuality, a fact which one will consider in the choice of patients. But one should be able to expect that this individuality is not a disturbing factor and that the training analysis should establish the necessary minimum in plasticity of character (6, p. 139).

From these considerations it would follow that, ideally, the prognosis for any one patient represents a realistic assessment of the severity of the disturbance reflecting clinical experience regarding the response of similar patients to psychotherapy.

While evidence has been presented (12) that a prognostic estimate *may* be based on the therapist's emotional reaction to the patient rather than on the clinical evidence, it does not follow that an unfavorable prognosis is *ipso facto* proof of a rejecting attitude on the therapist's part. It may well be a realistic assessment of the clinical picture.

It appears that the results obtained from client-centered therapists must be explained on different grounds. First, one of the most important conditions for constructive personality change postulated by Rogers (10) is the therapist's "unconditional positive regard" for the patient. This means a "prizing" of the person, the absence of *conditions* of accept-

ance and of a selective evaluating attitude. Unconditional positive regard appears closely related to empathy, another of Rogers' key postulates, which is defined as the therapist's "experiencing an accurate, empathic understanding of the client's awareness of his own experience" (10, p. 99).

Second, in explicitly rejecting the usefulness of diagnosis for psychotherapy, Rogers denies that the therapist's approach or operations should be influenced or modified by the type of patient or disturbance (10, p. 101). This would also imply that prognostic estimates, to the extent that they attempt to take into account the severity or chronicity of the patient's disturbance, are pointless.

Thus, it appears logical that the client-centered therapist who fully accepts the theory must (*a*) describe his attitude toward the patient as positive and (*b*) give a favorable prognosis. The ratings under discussion seem to reflect very clearly the respondents' theoretical orientation and are perhaps best explained on this basis.

Whether a conscious attitude toward the patient is a more or less individual expression of the therapist's emotional reaction or whether the attitude is an intrinsic part of the therapist's theoretical framework, or both, the more important question relates to the manner in which this attitude is expressed in the therapeutic situation. Some evidence (12) has been adduced that there is in fact a carryover. It remains to be demonstrated to what extent the therapist's initial attitude toward the patient influences the course and outcome of therapy, but there seems to be a very real possibility that the therapist making a negative evaluation and/or reacting negatively to the patient may unwittingly contribute to the realization of the events he diagnoses or predicts. Contrariwise, the therapist who brings a genuinely positive approach to bear upon the relationship may thereby promote a more favorable outcome.[9] Both psychoanalytic and client-centered theory emphasize that the person of the therapist, as revealed in this relationship, will in time serve as a new, more mature, and more desirable model of reality than past interpersonal relationships which have distorted the patient's perceptions of himself and others (13):

> What is curative in the [therapeutic] process is that in tending to reconstruct with the analyst that atmosphere which obtained in childhood, the patient actually achieves something new. He discovers that part of himself which had to be repressed at the time of the original experience. He can only do this in an interpersonal relationship with the analyst, which is suitable to such a rediscovery. . . . Thus, the transference phenomenon is used so that the patient will completely reexperience the original frames of reference, and himself within those frames, in a truly different relationship with the analyst, to the end that he can discover the invalidity of his conclusions about himself and others (7, p. 151).

Rogers puts it this way:

> In a minimal way, the client may perceive from the first the unconditional positive regard of the therapist for him. The perception is continually strengthened as he discovers that each facet of himself which is exposed—contradictions, weaknesses, strengths, abnormal feelings, tender feelings, vicious attitudes, antisocial behaviors, fears and despairs—are all met with equal positive regard, because each of these elements is a part of him, and he is prized unconditionally. The client then gradually takes in the experience of being loved—a love which is nonpossessive and nondemanding, warm but not oversolicitous. As he lives for a sequence of hours in this atmosphere, the experience has two primary results. It permits him to relax the tight defensive structure of his concept of himself and to admit into awareness and fully experience attitudes which previously he had found too threatening. . . . In the second place, the attitude of the therapist toward him is gradually internalized so that he can take the same attitude toward himself. He comes to prize himself, to feel that he *is* of value (9, p. 204).

Much remains to be learned about the therapist's personality and its contribution to the treatment process, but, as I have suggested elsewhere (11), this focus on the person of the therapist may be one of the most fruitful avenues for research in this area.

Summary

This investigation was designed to obtain empirical evidence on psychotherapists' performance in an initial interview for the purpose of studying the influence of relevant therapist variables.

Psychotherapists were asked to respond as vicarious interviewers to a sound film of an

[9] If it turns out that the therapist's personality—his dedication, faith, belief—is indeed a primary force in modern psychotherapy, painstaking research may still give a clearer picture of the components. (See in this connection [2].)

unrehearsed initial interview. Data were collected from 235 psychotherapists in several cities. The present paper, part of a larger project, was concerned with a comparison between 64 psychoanalytically oriented therapists and 14 therapists following Rogers' client-centered theory. All respondents were psychologists, and 8 Rogerians were staff members of the University of Chicago Counseling Center.

Statistical comparisons of questionnaire responses to a series of items dealing with diagnosis, prognosis, treatment plans, and attitudes disclosed a number of significant differences between the two samples. Rogerians were generally disinclined to make plans for treatment or to set up therapeutic goals. Their prognostic estimates were more favorable, and they rated their attitude toward the patient as more positive than analytically oriented therapists.

In discussing the implications of these findings, attention has been drawn to the interrelationships between therapist's attitude toward the patient and clinical evaluations, including diagnosis and prognosis. In view of the self-realizing character of one's expectations, the therapist's attitude toward the patient may have important effects upon the course and outcome of psychotherapy.

This investigation has stressed the great need for research on the therapist's contribution to the treatment process.

Received August 9, 1957.

References

1. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
2. Frank, J. D. Some effects of expectancy and influence in psychotherapy. In *Progress in psychotherapy.* Vol. 3. New York: Grune & Stratton, 1958. Pp. 27–43.
3. Fromm-Reichmann, Frieda. *Principles of intensive psychotherapy.* Chicago: Univer. Chicago Press, 1950.
4. Gitelson, M. The emotional position of the analyst in the psychoanalytic situation. *Int. J. Psychoanal.,* 1952, **33**, 1–10.
5. Orr, D. W. Transference and countertransference: A historical survey. *J. Amer. psychoanal. Ass.,* 1954, **2**, 621–670.
6. Reich, W. *Character-analysis.* (3rd ed.) New York: Orgone Inst. Press, 1949.
7. Rioch, Janet. The transference phenomenon in psychoanalytic therapy. *Psychiatry,* 1943, **6**, 147–156.
8. Rogers, C. R. Persons or science: A philosophical question. *Amer. Psychologist,* 1955, **10**, 267–278.
9. Rogers, C. R. Client-centered therapy: A current view. In Frieda Fromm-Reichmann & J. L. Moreno (Eds.), *Progress in psychotherapy 1956.* New York: Grune & Stratton, 1956. Pp. 199–209.
10. Rogers, C. R. The necessary and sufficient conditions of therapeutic personality change. *J. consult. Psychol.,* 1957, **21**, 95–103.
11. Strupp, H. H. A multidimensional system for analyzing psychotherapeutic techniques. *Psychiatry,* 1957, **20**, 293–306.
12. Strupp, H. H. The psychotherapist's contribution to the treatment process. *Behav. Sci.,* 1958, **3**, 34–67.
13. Strupp, H. H. Toward an analysis of the therapist's contribution to the treatment process. *Psychiatry,* in press.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

A Theoretical Framework for Nondirective Play Therapy: Concepts from Psychoanalysis and Learning Theory

Dell Lebo

Richmond Professional Institute

As Freud's systematic thinking progressed, aggression emerged from a position of relative obscurity in 1905 into one of decided prominence in 1930. At first, aggressive impulses were regarded as derivatives or components of sexuality (10). Later Freud contended that each instinct had its own power of becoming aggressive, deprivations being especially productive of aggressive responses (9).

Freud's formulation of the death instinct (6, 7) next led him to assert that a portion of that instinct, directed toward the outer world, was manifested as an aggressive instinct. However, it was not until 1930 (8) that a substantial body grew onto the theoretical framework of aggression. A desire for aggression was then said to be an independent, innate drive. This instinctual disposition varied in intensity from man to man. Libido and aggression were viewed as primary instincts entwined with strivings toward life and death (6).

The relationship between aggression and deprivation was experimentally examined by learning theorists and found to be relevant. Consequently, learning theory has given attention to aggression. At first it was assumed that frustration was a direct antecedent of aggression (2). As the investigations proceeded, it was realized that frustration could give rise to a variety of responses, one of which could be some form of aggression (17). Aggression was viewed as a secondary, rather than as a primary, drive because it did not seem to involve any biological changes in the organism. Instead, aggression was held to be a simple, learnable response, subject to modification. The selective agent in the learning of adult aggressive habits was reinforcement. The position of aggression in the hierarchy of responses to frustration depended upon whether past conditions were such as to reinforce aggression as a response to external cues. Despite a continued emphasis on the importance of learning in aggression, learning theorists no longer deny the role of innate factors (3).

There is an interesting parellel here. Two powerful psychological viewpoints seem to have gradually admitted aggression into more and more of their hypothetical realm. Such a similar change in two independent systems of thought may foreshadow the eventual importance of aggression in other psychotherapeutic theories. The theoretical development of aggression indicates it would be worthwhile to consider its possible importance to a psychotherapeutic theory that does not at present include it.

Aggression and Client-Centered Theory

One of the newest of psychological systems, the nondirective or client-centered approach, has not directly incorporated aggression. The fourth proposition of Roger's formulation of personality is concerned with "the tendency of the organism to behave defensively in the face of threat, to achieve the goal of self-maintenance even when the usual pathway to that goal is blocked" (22, p. 488). However, he included these tendencies, which might sound like aggression, with food assimilation

and a general striving toward maturation, the tendency to which all organic and psychological needs might be ascribed by nondirectivists. If the importance achieved by aggression in the two concepts previously discussed is a proper forecast, then it appears likely that it is through this proposition that aggression will enter nondirective personality theory.

Age and Aggression

In both psychoanalytic and learning theory the appearance and expression of aggression are closely dependent upon both maturational and experiential factors which follow a fairly close chronological sequence. To psychoanalysts, genetic factors must be considered in the establishment of libidinal development and characteristic defenses. It would be clinically incorrect for them to assume exclusively that an aggressive child is reacting constantly to external frustration and to exclude the considerations of inner sources such as ego disturbances and libidinal factors.

In analytic work with children, aggressive qualities reach their peak with the anal stage and become especially evident also during the Oedipal period (**5**). When the child reaches the stage in which his ego or superego disapproves of aggression turned outward, aggressive impulses will be directed to substitute objects, or modified by restriction, sublimation, and internalization (**12**).

The learning theorists explain these changes on another basis. Both schools would probably agree that the child is better able to handle instinctual drives by the acquisition of language. Psychoanalysts, however, see language integration as enabling ego mastery to proceed, while the learning theorists see it as an example of further learning which enables earlier, more diffuse, responses to be eliminated and more selective responses, such as verbal aggression, to be utilized. This verbal aggression, however, is subject to modification by reinforcement. As the child grows older and reactions with adults become important to him, his pattern of aggression is modified. Progression from infantile diffusion to directed attacks, to verbal specialization, to a marked reduction seems to be the behavioral pattern of aggression expression in normal children.

As a child grows chronologically, certain changes occur which influence the pattern of aggression. Whether these changes are called a developing superego or a modification through learning experiences, the position of aggression in the hierarchy of available responses does seem to change in an observable manner.

Other important behavioral changes take place from one age level to another. Gesell and Ilg have indicated that as a child grows older the "interior patterns of his private psychology" are changing as well (**11**, p. 12). Such changes have been a concern of psychoanalytically oriented play therapists who advocate different treatment approaches to children of different ages (**16**).

Age in Nondirective Play Therapy

Just as the concept of aggression has not been admitted into client-centered philosophy, so has the concept of personality changes with advancing age been lacking from formulations of the nondirective play therapy process. Perhaps the conceptualization of nondirective play therapy (**1**) excluded consideration of age differences because of the emphasis on *process* in nondirective therapy in play therapy: "Experiences with play therapy . . . add to the conviction that therapy is very definitely a process, a process which follows certain fundamental trends" (**21**, p. 170).

Nondirective play therapy, based directly upon the techniques of nondirective counseling, emphasizes the forces within each individual to find mature behavior more satisfying than immature behavior. "This type of therapy starts where the individual is and bases the process on the present configuration . . ." (**1**, p. 15). Since the individual is accepted where he *is* in client-centered therapy and the process begun, the child is accepted similarly where he is chronologically and the process of nondirective play therapy begun.

There may be a recognition that personality changes from age to age in the expression of the one basic nondirective need, namely, the striving for maturity. However, such an implied recognition of age differences

does not enter into the method of nondirective play therapy. The process is the same for all children admitted to the playroom regardless of age.

Studies of the Process of Nondirective Play Therapy

As has been shown, both age and aggression have been found to be factors of keen importance in other theories of psychotherapy, especially where children were concerned. Yet neither factor has been considered generally in studies of the process of nondirective play therapy. Such studies fall into two types, methodologically as well as chronologically distinguishable.

Studies of the earlier type involved examination of the process of nondirective play therapy as found in already available protocols. Later studies were objective investigations of the protocols of selected children.

Protocol-Centered Studies

Landisberg and Snyder (13) examined the protocols of four cases to determine what took place in nondirective play therapy. They found similarities in the process from case to case. Although the four children ranged in age from five to six years, the categories used to classify the verbal remarks had been developed for use with adults. Since the categories did not originate from children's verbalizations, their representativeness may have been forced and unrealistic. Furthermore, the narrow age range prevented inferences regarding the process of nondirective play therapy at different age levels.

The Procrustean forcing of children's statements into adult categories was obviated by Finke (4) who developed new categories based on an analysis of children's statements made in nondirective play therapy. She used protocols from six children ranging in age from five to eleven years. It was found that the children, though in therapy with different therapists, showed similar trends in their verbal expressiveness. Her study did not consider the possibility that just as the wide age discrepancies between adults and children might influence their verbalizations, so might children's categories vary significantly from one age level of maturity to another.

Seemingly the last study of this type was submitted for publication in 1952 (18). It was concerned with similarities in the pattern of emotional growth of normal and disturbed children. Extensive excerpts, arbitrarily selected and subjectively interpreted, revealed that "the emotional process and the changes in feeling tones evident as a child goes through play therapy . . . occur . . . not step by step, but in individually varying sequences" (18, pp. 97–98).

This study was nonobjective and resembled the suggestive investigations appearing in the persuasive or propagandistic period of research in nondirective play therapy (15). Nowhere was consideration given to age.

Child-Centered Studies

The first publication of the later type of investigation was apparently that of the present writer (14). Twenty children, aged 4, 6, 8, 10, and 12 years, equated for intelligence and social adjustment, were used as subjects. Maturation, as represented by chronological age, was found to account for different types of statements made in the nondirective play situation.

Similar studies by Moustakas (19, 20) have appeared since. One of them (19) studied the frequency and intensity of the expression of negative attitudes of nine well-adjusted and nine disturbed young children matched in various characteristics. At the end of four play therapy sessions, both groups expressed about the same types of negative attitudes.

Another example of subject-centered research concerned 10 four-year-old children with and without serious emotional problems (20). "Data obtained . . . showed more similar than divergent behavior in the two groups . . ." (20, p. 155). The most significant difference between the groups was that the maladjusted children were nonattentive. This difference was explained by the maladjusted children's noninteractive behavior, i.e., fantasy or activities that excluded the therapist. The degree of aggressiveness was not controlled. One wonders if the results would have been the same had aggression been considered.

The above mentioned studies represent all of the published work on the process of nondirective play therapy known to the present writer. While the studies are similar they are not directly comparable. In only four investigations does age seem to have been controlled. In three publications, age was regulated by restriction; only children within a narrow age range were studied: 5–6 years (13) and 4 years (19, 20) respectively. Hence, only one study clearly took age differences into account (14). This study alone found definite changes in the types of statements made by children in nondirective play therapy.

Only two of the studies clearly controlled intelligence (14, 19). Hence, differences in intellectual ability, long recognized as being of fundamental importance in speech habits (*vide* the items of the Stanford-Binet Intelligence Test), have not been considered in many studies of the verbalizations made in nondirective play therapy. This lack would seem to be of sufficient magnitude to vitiate investigations that have failed to control for it. None of the studies considered the possibility of differences arising in verbalizations as related to aggressiveness or lack of aggressiveness in children.

The empirical progress made in psychoanalysis and learning theory has been anticipated by the construction of adequate theory. Further developments in nondirective play therapy would seem to require suitable theory. Without such a theoretical framework, nondirective play therapy may be in danger of becoming immeshed in the "regularity" of its process, until, like introspectionism, it becomes process bound and unproductive. A good theory is said to be economical in permitting many observations to be subsumed under a single proposition. Theories concerning changes ascribable in aggression or age would seem to be both economical and meaningful.

Summary

Psychotherapeutic processes may vary considerably as a function of both the age of the subject and the dynamic processes characterizing his psychopathology. In childhood, the relationship between age and aggression is especially close. Nondirective play therapy does not formally consider the influence of aggression or age in its method. As a result, studies of the process of nondirective play therapy generally indicate that all children undergo the same process. It would seem, however, that if age and aggression were evaluated properly, the course of nondirective play therapy would vary with the age of the child and with the character of the psychodynamic processes, e.g., aggression, associated with or defining his problem. Insistence on the sameness of the process, and consequently of the method, may only result in the tautological death of productive work in nondirective play therapy.

These assertions were supported by inferences drawn from the history of Freudian and learning theory and by data from studies of the process of nondirective play therapy.

Received August 12, 1957.

References

1. Axline, Virginia M. *Play therapy*. Boston: Houghton Mifflin, 1947.
2. Dollard, J., Doob, L., Miller, N. E., Mowrer, O. H., & Sears, R. R. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.
3. Dollard, J., & Miller N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
4. Finke, Helene. Changes in the expression of emotionalized attitudes in six cases of play therapy. Unpublished master's thesis, Univer. of Chicago, 1947.
5. Freud, Anna. Notes on aggression. In S. Lorand (Ed.), *The yearbook of psychoanalysis*. Vol. 6. New York: International Universities Press, 1951. Pp. 145–154.
6. Freud, S. *Beyond the pleasure principle*. London: Hogarth Press, 1922.
7. Freud, S. *The ego and the id*. London: Hogarth Press, 1927.
8. Freud, S. *Civilization and its discontents*. London: Hogarth Press, 1930.
9. Freud, S. *A general introduction to psychoanalysis*. New York: Liveright, 1935.
10. Freud, S. Three contributions to the theory of sex. In A. A. Brill (Ed.), *The basic writings of Sigmund Freud*. New York: Modern Library, 1938. Pp. 553–629.
11. Gesell, A., & Ilg, Frances L. *The child from five to ten*. New York: Harper, 1946.
12. Hartmann, H., Kris, E., & Lowenstein, R. M. Notes on the theory of aggression. In Phyllis Greenacre *et al.* (Eds.), *The psychoanalytic study of the child*. Vol. 3–4. New York: International Universities Press, 1949. Pp. 9–36.

13. Landisberg, Selma, & Snyder, W. U. Nondirective play therapy. *J. clin. Psychol.*, 1946, **2**, 203–213.
14. Lebo, D. The relationship of response categories in play therapy to chronological age. *J. Child Psychiat.*, 1952, **2**, 330–336.
15. Lebo, D. The present status of research on nondirective play therapy. *J. consult. Psychol.*, 1953, **17**, 177–183.
16. Lebo, D. Age and suitability for nondirective play therapy. *J. genet. Psychol.*, 1956, **89**, 231–238.
17. Miller, N. E., Sears, R. R., Mowrer, O. H., Doob, C. W., & Dollard, J. The frustration-aggression hypothesis. *Psychol. Rev.*, 1941, **48**, 337–342.
18. Moustakas, C. E. Emotional adjustment and the play therapy process. *J. genet. Psychol.*, 1955, **86**, 79–99.
19. Moustakas, C. E. The frequency and intensity of negative attitudes expressed in play therapy: A comparison of well-adjusted and disturbed young children. *J. genet. Psychol.*, 1955, **86**, 309–325.
20. Moustakas, C. E., & Schalock, H. D. An analysis of therapist-child interaction in play therapy. *Child Develpm.*, 1955, **26**, 143–157.
21. Rogers, C. R. *Counseling and psychotherapy*. Boston: Houghton Mifflin, 1942.
22. Rogers, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Personality Patterns of Patients Before and After Application for Psychotherapy[1]

Norman S. Greenfield
University of Wisconsin Medical School

A previous study (1) utilized the MMPIs of 132 male college students who later at some time in their four-year college course sought psychotherapy at a student health psychiatric clinic, with the aim of testing the hypothesis that promptness of referral is positively related to the degree of anxiety, the extent of internalization of problems, the amount of discomfort, and the severity of pathology. The results were surprising in that none of the variables proved to be significantly related to the number of months which intervened between the time of taking the test and appearance at the clinic. Thus, in the search for precise motivational variables influencing utilization of psychotherapeutic services, it was impossible to verify even such simple "truths" which have so long received consensual validation. It was concluded that for the population studied, at least, these measures were not indices of therapeutic readiness.

Having demonstrated an inability to predict which students would eventually become patients and which would not in a setting wherein psychotherapy was equally available to all students at no cost to the patient, the question was then raised as to what personality factors, or more precisely what changes in personality, influenced the decision to apply for treatment. Again, logic and clinical experience led us to assume that, other factors being equal, a person seeks psychotherapy at a time when he is feeling worse, or suffering more than when he did not seek psychotherapy.

The routine administration of the MMPI[2] to entering students made possible an examination of self-controlled profiles of 31 students who were given a second MMPI at the clinic. The *Ss* were 20 males and 11 females whose mean age was 19 years. The mean number of months intervening between first and second test administrations was 10.97, with a range from 1 to 26 months.

Clinically, the presenting complaints of the patients as well as their psychiatric diagnoses reflected the usual range of patient material encountered at the clinic.

None of the differences between the scores of the two administrations approached significance. Both in pattern and amplitude pre- and posttests were almost identical.

The voluminous body of literature utilizing the MMPI to demonstrate differences between groups and, even more sensitively, changes in scores following psychotherapy suggests caution in cavalierly discarding these data and turning to the parsimonious challenge of the validity of the technique. In any case, these results furnish some food for thought.

Brief Report.
Received January 30, 1958.

Reference

1. Greenfield, N. S., & Fey, W. F. Factors influencing utilization of psychotherapeutic services in male college students. *J. clin. Psychol.*, 1956, 12, 276–279.

[1] An extended report of this study may be obtained without charge from Norman S. Greenfield, Department of Psychiatry, University Hospitals, Madison 6, Wisconsin, or for a fee from the American Documentation Institute. Order Document No. 5584, remitting $1.25 for microfilm or $1.25 for photocopies.

[2] Some of the data for this study were made available through the generous cooperation of Lewis E. Drake.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Diagnostic Validity of the Ames "Danger Signals"

Elise Elkins[1]
Illinois Institute for Juvenile Research

The application of the Rorschach to very young subjects has resulted in an accumulation of studies indicating that children normally give responses that are pathological in adults (10, 13). The publication of *Child Rorschach Responses* by Ames and her associates (1) represented a milestone in that developmental norms were presented systematically at yearly intervals for a carefully studied group of children ten years of age and under. Toward the end of this volume, Ames presented a list of "danger signals" which, even within the developmental context of a child's Rorschach, might still be considered indicative of pathology. The present study was designed to test whether these "danger signals" had sufficient diagnostic validity to differentiate between a clinic and a normal population. As part of a study of normal eight-, nine-, and ten-year-olds undertaken by the Institute for Juvenile Research, 54 Chicago school children described as well-adjusted by their teachers were given a complete diagnostic evaluation. For the purposes of this study, an effort was made to match each of these nonclinic cases with a child who had been referred to the Institute as being in need of child guidance services. The present report deals with a comparison of the Rorschach records of 40 nondisturbed children with the records of their matched controls.

Subjects

The *S*s were 40 normal children (27 boys and 13 girls) and their controls who were matched for sex, age, and IQ. Matching was in most cases within three months of age and five points of IQ. The matching for IQ was only approximate. All of the nondisturbed *S*s were given Form L of the Binet by the same examiner, while their matched controls were given a variety of other tests, both group and individual. In all likelihood, the clinic group is slightly more intelligent than the nondisturbed sample, since the clinic children were often matched on the basis of tests scores that are known to run lower than Binet ratings. This direction of a possible difference was considered safest so that there would be no question of the clinic children's looking "sick" only because they were merely less intelligent. Actually, both the clinic and the nondisturbed children taken together constitute a highly homogeneous sample of children of superior intelligence. In this respect, both groups are comparable to the normative sample used by Ames (1, p. 25).

Although no effort was made to match for socioeconomic status, the matching for IQ resulted in a clinic and a nondisturbed sample which were quite comparable in this regard. The socioeconomic level of the *S*s was evaluated by means of the scale of paternal occupations described by Goodenough (6, pp. 501–512). Three clinical psychologists classified the 80 *S*s. In 39 of the 80 cases, all three judges agreed on the classification. In an additional 31, one judge differed by one category from the other two. The *S* was then assigned to the category agreed upon by two out of three judges. In the nine cases with ratings spread over three categories, the *S* was assigned to the middle category. Both the clinic and the nondisturbed group are slightly below the Ames sample in regard to socioeconomic status: 93% of the Ames sample (1, p. 25) and 59% of the combined

[1] The author would like to express her gratitude to Charlotte Altman who tested the entire nondisturbed sample; to Eugene Levitt who gave helpful suggestions; and to Irene Goldblatt, Jeanne Hershenson, and Victor Lebow who served as scorers and raters.

clinic and normal groups of this study fall into Categories I through III on the scale of paternal occupations. The atypically high socioeconomic status of the Ames sample in comparison to the general population raises the possibility of bias in the Ames normative data. In fact, Fiedler and Stone (**5**, p. 279) found striking differences between their sample of undisturbed children of low socioeconomic status and the Ames sample of high socioeconomic status. However, in this study, any possible effects of bias in the Ames norms are controlled by the matched-pairs design.

Procedure

All 40 Rorschachs of the nondisturbed children were administered and scored by a psychologist with 15 years' experience and certification by the American Board of Examiners in Psychology. The 40 Rorschachs of the matched controls were administered and scored by 15 psychologists of varying levels of experience. Because of the extreme difficulty of obtaining enough clinic cases to match with a group of normals of such high intelligence, it was not possible to control for experience or identity of the examiner. Since arrangements could not be made to have all 80 Rorschachs retyped without identifying information, this study was limited to objective signs whose presence or absence could be evaluated merely by a check of the scoring sheet.

Table 1

Descriptive Information About Subjects Used in This Study

Variable	Nondisturbed Group $N = 40$	Clinic Group $N = 40$
Age		
Range	8–4—10–0	8–3—9–11
Mean	108.38 mos.	108.15 mos.
SD	5.50 mos.	5.40 mos.
Sex		
Male	27	27
Female	13	13
IQ		
Range	98–145	100–155
Mean	120.15	120.30
SD	12.52	14.06
Socioeconomic status		
I. Professional	6	8
II. Semiprofessional and managerial	6	5
III. Clerical, skilled trades, retail business	11	11
IV. Semiskilled, minor clerical	11	9
V. Slightly skilled	4	4
VI. Day laborers	2	3

Of the 13 "danger signals" listed by Ames and her associates, 4 were dropped because they involved subjective judgments. The remaining 9 were raised to a total of 13 because of the necessity of treating the parts of four complex signs as separate signs (e.g., Sign 10: *A*% over 60 and no human responses). The 13 signs used in this study are listed below with the numbering used by Ames (**1**, p. 283). Qualifications introduced by the author are given in parentheses.

1. Preponderance of tree or leaf responses (Preponderance defined as over 50% of *R*)
2. *F*% exceeds 75
3. More than one pure color response

5A. More than one response involving differentiated shading

5B. More than one response involving diffuse, dysphoric shading, i.e., *Clob*

6. More than one inanimate movement response
8. W% less than 40

9A. *Dd*% over 15. *Dd*% equals *Dd*, *S*, and *Do* responses over total *R*

9B. More than one *X* response

9C. More than one *S* response

10A. *A*% over 60

10B. No human response (i.e., no *H*: *Hd* not counted)

11. *R* less than 10

The evaluation of the 13 signs was done by the author. A slight element of subjective judgment could not be avoided in that some rescoring had to be done before even the objective Ames signs could be checked. Of the 80 protocols used in the study, 70 were scored by the Beck system (**2**), while the Ames system is closer to the Klopfer method (**9**). Fortunately, it has been common practice at the Institute to make marginal notations of *m* and *FM*. However, the elimination of separate *D* scorings for details covered in a *W* and the division of shading scores into *Clob* (diffuse, dysphoric) and *F(C)* (differenti-

ated) also had to be done. The author prepared a detailed instruction sheet indicating the differences between the Ames and the usual clinic scoring in regard to those variables that would be involved in an evaluation of the Ames danger signals. The scoring record for each of the 80 protocols was then transferred to a new scoring sheet, with each response being scored only for the relevant variables which are as follows:

Area (*W, D, Dd, Do* response, *S*)
Major determinant (*M, m, FM, C,* either *FC* or *CF,* differentiated shading, diffuse shading with dysphoric element, *F*)
Content (*H, A* or *Ad,* tree or leaf)

No differentiations other than those listed were made. For example, no effort was made to differentiate between *FC* and *CF*. The Ames signs were then checked by counting the various types of responses on these scoring sheets. Tabulation of the Ames signs was, of course, purely clerical once the scoring sheets had been made out.

As for the possibility of the introduction of bias in the rescoring, the author can state only that there seems to be satisfactory evidence in the literature that mere scoring of Rorschach responses has sufficient interexaminer reliability to warrant being regarded as an objective procedure (**7, 8**). However, as a reliability check, one-fourth of the protocols (10 normals and 10 clinic cases) were scored by a colleague as well as by the author. The percentage of agreement on the scoring of location varied from 71 to 93 on the 20 records, with a mean of 81 and an average deviation of 5.25 percentage points. The percentage of agreement on the scoring of major determinant varied from 65 to 100 on the 20 records, with a mean of 84 and an average deviation of 7.05 percentage points. The percentage of agreement on the scoring of content varied from 77 to 100 on the 20 records, with a mean of 91 and an average deviation of 6.10 percentage points. There was perfect agreement on content scoring for a third of the records.

Results

Total Scores

The total scores on the Ames list of danger signals were compared for the clinic and the nondisturbed groups by means of a variation of McNemar's Formula 26 for the evaluation of the difference between two correlated means (**12**, p. 74). The results are presented in Table 2. The clinic group had significantly fewer danger or maladjustment signs than the nondisturbed group. Before investigating the direction of the discrimination done by each individual Ames sign, the author compared the clinic and the nondisturbed groups by means of 13 of the 17 Davidson adjustment signs (**4**). The two signs involving color shock and shading shock respectively were omitted because they involve subjective judgment. The two signs involving *FM* and *F%* respectively were omitted because the records were scored by the Beck system. The Davidson signs were evaluated on the basis of the scoring done by the original examiners without any rescoring by the author. The results of this comparison are also presented in Table 2. In contrast to the Ames signs, the Davidson signs show an absolute difference in the expected direction; however, the difference is so small that there is a high probability that it could be a chance deviation from a true difference of zero ($p > 60$). Although all statistical tests were done on the basis of the 13 signs actually counted for both the nondisturbed and the clinic groups, the mean based on 13 signs was prorated to obtain the estimated mean had all 17 signs been used. The prorated means are also shown in Table 2. An estimate of the accuracy of the prorating

Table 2

Total Scores on Ames "Danger Signals" and Davidson Adjustment Signs in Matched Pairs of Clinic and Nondisturbed Subjects

	Nondisturbed Group	Clinic Group	*CR*
Ames "Danger Signals"			
Mean	3.00	2.20	2.25*
SD	1.55	1.75	
Davidson Adjustment Signs			
Mean	7.10	6.58	.57
SD	2.27	2.84	
Prorated mean	9.26	8.60	

* Significant at .05 level.

is provided by some additional data available for the nondisturbed sample only. The original examiner had evaluated the group on the basis of all 17 signs and obtained a mean of 9.3. The prorated mean obtained by extrapolating from the mean of 7.1 for 13 signs is also 9.3. The prorated means of 9.3 and 8.6 for the nondisturbed and the clinic groups respectively are consistent with the results of previous studies which show slight differences between normal and disturbed groups with considerable overlapping of the distributions of scores. In the studies reviewed by Davidson, means of 9.05 to 11.34 were reported for better-adjusted groups and means of 6.85 to 8.17 for less well-adjusted groups (4).

Individual Signs

The Davidson signs (4) did not differentiate the nondisturbed from the clinic patients in the present study, and the Ames danger signals discriminated in the "wrong" direction when evaluated in terms of total scores. Because of the possibility that individual Ames signs discriminated in different directions, the author carried out a detailed analysis, sign by sign. The results of this analysis are presented in Table 3.

Table 3

Incidence of Ames "Danger Signals" in Matched Pairs of Clinic and Nondisturbed Subjects

Danger signal	Both of pair	Neither of pair	Only clinic	Only Nondisturbed	Chi square
1A Preponderance of tree and leaf *R*	0	40	0	0	0
2 *F*% > 75	1	26	4	9	1.23
3 Pure *C* > 1	0	38	2	0	.50
5A Differentiated shading > 1	3	21	6	10	.56
5B Diffuse, dysphoric shading > 1	0	36	3	1	.25
6 *m* > 1	0	36	4	0	2.25
8 *W*% < 40	17	3	6	14	2.45
9A *Dd*% > 15	12	12	4	12	3.06
9B *X* responses > 1	0	33	5	2	.57
9C *S* > 1	2	33	2	3	0
10A *A*% > 60	2	16	5	17	5.50*
10B No *H*	4	24	4	8	.75
11 *R* < 10	0	39	0	1	0

* Significant at .05 level.

In the case of each sign, the number of instances of occurrence in the nondisturbed group and lack in the clinic cases were compared with the number of instances of occurrence in the clinic group and lack in the nondisturbed cases by means of McNemar's formula for chi square for dependent samples (12, pp. 204–206). The use of McNemar's formula permits one to evaluate only those two cells of the fourfold table which represent the difference critical for one's hypothesis. In this case, only the third and fourth columns of Table 3 were compared. The Yates correction for continuity was used in all computations.

The data in Table 3 suggest first of all that most of the Ames danger signals are not useful for differentiating between clinic and nondisturbed samples of eight- to ten-year-olds because nearly half of the signs are either characteristically present in a sizable proportion of both normal and clinic cases or else are almost never found in the records of either. Thus in 42% of the matched pairs, both the clinic and the nondisturbed member have a *W*% below 40. In 100% of the matched pairs, neither member has a preponderance of tree and leaf responses. In 90% of the matched pairs, neither member has more than one response involving diffuse, dysphoric shading. The only "danger sign" which has differentiating power that reaches statistical significance is Sign 10A (*A*% over 60). This sign discriminates in an unexpected direction: Of 22 matched pairs whose members differ on this sign, three times as many nondisturbed as clinic cases have the high *A*%. The most promising signs which discriminate in the expected direction are Signs 3 (pure $C > 1$), 5B (more than one instance of diffuse, dysphoric shading) and 6 ($m > 1$). These signs, like contamination for schizophrenia, are apparently diagnostic when they occur, but have such limited incidence as to be of little use for routine screening purposes.

Ranges on Determinants Involved in the Ames Danger Signals

The Ames signs, in contrast to the Davidson signs, make less allowance for ranges of scores with an optimum somewhere in the middle. Therefore, at the risk of inflating

Table 4

Comparison of Matched Pairs of Clinic and Nondisturbed Subjects in Regard to Scores Below the 25th Percentile of the Ames Norms

Determinant and level necessary to fall below 25th percentile	Both of pair below	Neither of pair below	Only clinic below	Only nondisturbed below	Chi square
F% below 52	4	20	8	8	0
No pure *C*	31	0	4	5	0
Differentiated shading zero	10	15	5	10	1.07
Diffuse, dysphoric shading zero	17	1	3	19	10.23**
m of zero	20	4	3	13	5.06*
W% < 28	11	8	7	14	1.71
Dd% zero	1	31	6	2	1.12
A% < 40	3	24	6	7	0
R < 13	0	33	2	5	.57

* Significant at .05 level.
** Significant at .01 level.

probabilities, one further analysis of the data was performed. The matched pairs of clinic and nondisturbed cases were compared in regard to the determinants underlying the danger signals in two ways: (*a*) frequency at or below the 25th percentile [2] of the Ames norms and (*b*) frequency above the 75th percentile of the Ames norms (1, pp. 54–56). These findings are presented in Tables 4 and 5.

Table 4 indicates that in the case of those matched pairs whose members differ in regard to being below the 25th Ames percentile, there are significantly more pairs whose nondisturbed member is below the 25th percentile for number of *m* and for number of responses involving diffuse, dysphoric shading. (As before, only Columns 3 and 4 were compared by means of McNemar's formula for chi square for dependent samples.) Table 5 indicates that in the case of those matched pairs whose members differ in regard to being above the 75th Ames percentile, there are significantly more pairs whose nondisturbed member is above the 75th percentile for *A*% and significantly more pairs whose clinic member is above the 75th percentile for number of responses involving diffuse, dysphoric shading.

A comparison of Tables 4 and 5 with Table 3 lends no support to the author's suspicion that the Ames danger signals would be improved by the introduction of middle ranges of scores on various determinants. Three of the four significant differences in the latter two tables result from a change in the cutting score for variables which appear to have a linear relationship to good adjustment. Thus, a complete lack of *m* and of dysphoric shading, rather than the presence of a moderate amount, differentiates the clinic from the nondisturbed sample. In the case of the fourth significant difference on *A*%, the data do not reveal whether there is any point at which the relationship between health and *A*% ceases to be linear in regard to the groups used in this study. The highest *A*% of 93 was found in the nondisturbed group, in addition to the fact that more of the undisturbed group exceeded the Ames 75th percentile (*A*% of 58).

Discussion of Results

The Ames danger signals offer no advantages over the other sign lists developed without special attention to the younger age groups and the developmental sequences particular to them. Neither are the Ames danger

Table 5

Comparison of Matched Pairs of Clinic and Nondisturbed Subjects in Regard to Scores Above the 75th Percentile of the Ames Norms

Determinant and level necessary to exceed 75th percentile	Both of pair above	Neither of pair above	Only clinic above	Only nondisturbed above	Chi square
F% above 78	2	26	3	9	2.08
Pure *C* > 1	0	38	2	0	.50
Differentiated shading over 1	3	21	6	10	.56
Diffuse, dysphoric shading > zero	1	18	18	3	9.33**
m > 1	0	37	3	0	1.33
W% > 52	1	27	6	6	0
Dd% > 16	11	12	4	13	3.76
A% > 58	4	15	4	17	6.86**
R > 23	6	16	11	7	.50

** Significant at .01 level.

[2] In the case of those signs with zero as the score for the 25th percentile, comparisons had to be made in regard to comparative frequencies at the 25th percentile since, of course, there could be no frequencies below zero.

signals any worse for screening purposes. The fact that several of the Ames signs discriminate in a direction opposite to that expected has been found to be true of comparable sign lists such as the Davidson adjustment signs. Brockway, Gleser, and Ulett (3), for example, found that 8 of the 17 Davidson signs discriminated in a reverse direction when used on a population of adult males.

One of the studies most comparable to the one described in this paper was carried out at the same clinic by another investigator. Levitt (11) compared 39 clinic cases, including 4 used by the present author, with the normative sample of Thetford, Molish, and Beck in regard to alleged Rorschach anxiety indices. He found that only *Sum Y* discriminated in the expected direction at better than the .05 level of confidence. The discriminating power of *Y* was improved by separating the differentiated shading responses (*FY*) from those with less of a form element (*YF*). Similarly, in the study by the present author, the best discrimination was obtained by the use of *Clob* responses only, the latter being roughly comparable to the *YF* responses of Levitt's study. However, the sum of all shading responses had some discriminating power. When the 40 matched pairs of clinic and nondisturbed subjects were compared in regard to total number of shading responses, significantly more pairs were found with the normal subject at or below the Ames 25th percentile [3] of zero shading responses (chi square 5.26 and $p < .05$). When a similar comparison was made in regard to scores above the Ames 75th percentile, there was a statistically insignificant trend for more clinic than normal members of matched pairs to be above the Ames 75th percentile for total number of shading responses. That is, complete lack of shading responses is significantly associated with membership in the nondisturbed group, while having two or more shading responses is less discriminating. Once shading responses are present, it is necessary to know the type before meaningful discriminations can be made between groups.

In general, those signs and determinants which discriminate in any way between the clinic and the disturbed group seem to have to do with tension (*m*, shading responses) and the stereotypy-conformity dimension of behavior (*A%*). Both the clinic and the nondisturbed samples are children of latency age. The nondisturbed sample appears to be distinguished mainly by a consolidation of defenses. Apparently, deviation in the direction of more control than the normative Ames sample is not pathological at this age, whereas deviation in the direction of less control identifies the patient population. Although in the adult, magnitude of deviation may be of major importance, it is possible that in the child the direction of the deviation in terms of the child's developmental level is equally important. The nondisturbed group in this study is exaggeratedly like the conforming, controlled model child that is the societal norm for the latency period. The fact that the nondisturbed sample was chosen by school teachers undoubtedly increased the probability that a particular variety of the societal ideal of normal latency adjustment would constitute the sample. Perhaps in the expansive period of adolescence the deviations in the direction of overcontrol are the more pathognomonic than similar quantitative deviations in the direction of impulsivity.

[3] The Ames percentile for total number of shading responses was estimated by adding the number of differentiated and diffuse shading responses at the given percentile. Thus the 25th percentile for all shading responses was estimated as zero, since zero is the 25th percentile for both *F(C)* and *Clob*. The 75th percentile was estimated as 1 since the 75th percentile for *F(C)* is zero and for *Clob* is 1.

Summary

Forty children described as well-adjusted by their teachers were matched for sex, age, and IQ with children referred for child guidance services to a large community clinic. Each group contained 13 girls and 27 boys. The average age of both groups was a little over nine years, and the average IQ of both groups was a little over 120. The matched pairs were compared by means of the Rorschach in the following ways: (*a*) total number of objectively scored Ames danger signals out of a list of 13, (*b*) total number of objectively scored Davidson adjustment signs out of a list of 13, (*c*) presence or absence of individual Ames danger signals, (*d*) compara-

tive frequencies above the Ames 75th percentile on determinants underlying the Ames danger signals, and (*e*) comparative frequencies below the Ames 25th percentile for determinants underlying the Ames danger signals.

The nondisturbed group was found to have a significantly higher total of Ames danger signals. The difference between the two groups on the Davidson adjustment signs did not reach statistical significance, but was in the expected direction. When the individual Ames signs were evaluated separately, only *A%* above 60 was found to discriminate at a statistically significant level. The difference was in the reverse of the expected direction, i.e., the nondisturbed group had more members with the high *A%*. When comparative frequencies above and below the Ames 75th and 25th percentiles respectively were evaluated, only the following differences between the clinic and the nondisturbed group were found to be statistically significant ($p < .05$): fewer clinic children with no *m*, fewer clinic children with *A%* over 58, more clinic children with responses involving diffuse, dysphoric shading, and more clinic children having two or more of the latter type of response.

The author concluded that the Ames danger signals do not have sufficient diagnostic validity to make them useful for differentiating between maladjusted and nondisturbed samples of eight- to ten-year-olds. Nearly half of the signs are either characteristically present in a sizable proportion of both normal and clinic cases or else are almost never found in the records of either. Several of the "danger" signs actually tend to be found *more* frequently in normal records. However, the developmental frame of reference underlying the signs was still seen as useful in interpreting the findings of the study.

Received July 22, 1957.

References

1. Ames, Louise B., Learned, Janet, Metraux, Ruth W., & Walker, R. N. *Child Rorschach responses: Developmental trends from two to ten years*. New York: Hoeber-Harper, 1952.
2. Beck, S. J. *Rorschach's test*. Vol. I. *Basic processes*. New York: Grune & Stratton, 1949.
3. Brockway, Ann L., Gleser, Goldine, & Ulett, G. A. Rorschach concepts of normality. *J. consult. Psychol.*, 1954, **18**, 259–265.
4. Davidson, Helen H. A measure of adjustment obtained from the Rorschach protocol. *J. proj. Tech.*, 1950, **14**, 31–38.
5. Fiedler, M. F., & Stone, L. J. Rorschachs of selected groups of children and published norms: II. The effect of socio-economic status on Rorschach performance. *J. proj. Tech.*, 1956, **20**, 276–280.
6. Goodenough, Florence L., & Anderson, J. E. *Experimental child study*. New York: Century, 1931.
7. Hertz, Marguerite. Scoring the Rorschach inkblot test. *J. genet. Psychol.*, 1938, **52**, 15–64.
8. Hertz, Marguerite. On the standardization of the Rorschach method. *Rorschach Res. Exch.*, 1939, **3**, 120–133.
9. Klopfer, B., & Kelley, D. *The Rorschach technique*. Yonkers, N. Y.: World Book, 1946.
10. Klopfer, F., & Margulies, Helen. Rorschach reactions in early childhood. *Rorschach Res. Exch.*, 1941, **5**, 1–23.
11. Levitt, E. E. Alleged Rorschach anxiety indices in children. *J. proj. Tech.*, in press.
12. McNemar, Q. *Psychological statistics*. New York: John Wiley, 1949.
13. Paulsen, Alma. Rorschachs of school beginners. *Rorschach Res. Exch.*, 1941, **5**, 24–29.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

An Investigation of Determinants Inducing Individuals to Seek Personal Counseling[1,2]

James S. Terwilliger and Fred E. Fiedler
University of Illinois

This study attempted to identify some of the determinants which lead individuals to seek therapeutic help. Very few situations permit an adequate study of this problem. Generally the counselor only sees people who feel the need for psychotherapy, while many others, equally maladjusted by various criteria, choose, for various reasons, not to consult a therapist. It is, therefore, of considerable theoretical and practical importance to determine how those who seek counseling differ in important respects from those who do not.

Such a study was possible at the University of Illinois where the Student Counseling Service deals with students' educational, vocational, and personality problems. Students are urged to consult counselors if they, at any time, feel the need for professional help. Counseling services are provided without charge.

Sample. The subjects (*Ss*) were male University freshmen who lived in temporary barracks which normally housed 13 to 14 men. *Ss* are assigned to the barracks in the order in which their application is received.

On the basis of counseling service records, we identified a "counseling bureau group" (CBG) consisting of *Ss* who had visited the Counseling Bureau two or more times for personality problems ($N = 10$). Those who had never consulted the bureau were placed into the control group ($N = 87$).

Tests. A number of personality and adjustment measures were obtained from each *S*. These included an abbreviated Taylor Manifest Anxiety Scale, a self-esteem score, a self-ideal self discrepancy (or self-satisfaction) score. Also obtained were sociometric status indices to provide a measure of the individual's acceptability to his peers, and "esteem by others" scores from *S*'s fellow students who lived in the same barracks. This latter score indicates *S*'s adjustment as evaluated by others. In addition, we also computed a Grade Point Average, adjusted for *S*'s performance in high school.

Results. Taylor anxiety scores, self-satisfaction, and self-esteem scores of *Ss* who consulted therapists differed significantly in the maladjusted direction from those who did not consult a therapist. However, differences in other scores were not significant.

Discussion. As is to be expected, anxious and less self-satisfied individuals are more likely to seek therapeutic help than are those who do not express such feelings. In contrast, however, these *Ss* do not differ from the control group on objective or outside indices. *Ss* who consulted therapists were as well liked as others, they were esteemed as highly, and they tended to have slightly better grades. These findings suggest that the client is brought into therapy solely by subjective feelings of discomfort with himself rather than by "reality" factors.

Brief Report.
Received February 20, 1958.

[1] This study was conducted under Contract DA-49-007-MD-569 between the Office of the Surgeon General, Department of the Army, and the University of Illinois.

[2] An extended report of this study may be obtained without charge from Fred E. Fiedler, 427 Gregory Hall, University of Illinois, Urbana, Illinois, or for a fee from the American Documentation Institute. Order Document No. 5582, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Vocabulary as a Basis for Estimating Intelligence from the Rorschach[1]

Thomas R. Trier
University of California, Berkeley

An estimate of intelligence is usually included in the formulations made from the Rorschach. It can be particularly helpful when time limitations or other considerations render it impractical or impossible to administer an intelligence test. Beck (2) and Klopfer and Kelley (3) are among those who have tried to make more explicit the variables which should be considered in estimating intellectual level. Beck states that intelligence is related to *Z*, *A%*, *F + %*, *Ap*, and number and quality of *M* responses. Klopfer and Kelley suggest that an estimate of intelligence should be based on number and quality of *W*, *M*, and *O* responses, form accuracy level, variety of content, and succession of responses. Research seeking to determine whether the postulated relationships between Rorschach variables and IQ do in fact exist has been rather unsuccessful. Usually, when low, significant correlations between Rorschach factors and IQ have been found, they have failed to hold up under cross-validation. The results reported by Armitage *et al.* (1) are typical in this respect, but are particularly worthy of mention because of the large number of cases employed. The authors computed the correlations between 19 Rorschach variables and Wechsler-Bellevue IQ for 503 patients. Low correlations, significant at the .01 level, were obtained for 13 of the variables. A multiple regression equation, based on the six variables showing the highest correlation with IQ, was computed. When this equation was applied to a new sample of 207 patients, it failed to improve the accuracy of estimates beyond chance. In contrast, judges were able to make fairly accurate estimates of intelligence when exposed to the total protocols. The authors suggest the possibility that the judges were responding to the quality of the vocabulary and the quality of the perceptual organization in making their judgments.

The present study is an attempt to determine whether the vocabulary in Rorschach protocols is a variable that can be profitably utilized in estimating intelligence. Wechsler's statement, ". . . the size of a man's vocabulary . . . is an excellent measure of his general intelligence" (6, p. 98), has been generally accepted. Vocabulary subtests are included in many intelligence tests. In both the Stanford-Binet (4) and Wechsler-Bellevue (6), the vocabulary subtest correlates higher with total IQ than do any of the other subtests. If the extensiveness of a person's vocabulary can be judged from the Rorschach with some accuracy, this measure might be useful in estimating his intelligence.

Procedure

Three groups, each consisting of four judges, were asked to make numerical estimates of Wechsler-Bellevue IQ. Judges in Group I were given 16 Rorschach protocols. They were asked to list the seven "most sophisticated" words from each protocol and to base their IQ estimates on these word lists. Judges in Group II were each given one of the four sheets of word lists constructed by the judges of Group I. They were informed how the lists were derived and were asked to make IQ esti-

[1] This research was performed while the writer was a psychology trainee at the Veterans Administration Hospital, San Francisco, California. He wishes to express his appreciation to David A. Rodgers for his valuable assistance in the design and analysis of the experiment and in the preparation of the article.

mates using only the word lists. They differed from Group I in having no access to the total protocol. Group III judges were handed the sixteen protocols and asked to estimate intelligence from them. In contrast to the others, Group III judges were not given specific instructions on what to base their judgments.

In addition, an attempt was made to measure the extensiveness of vocabulary in a more objective fashion. The four word lists constructed by Group I judges for each protocol were pooled. While these word lists overlapped considerably, pooling revealed that from 10 to 20 different words had been selected as "most sophisticated" for each protocol by the four judges, with a mean of 14.7. These words were looked up in Thorndike and Lorge's *The Teacher's Word Book of 30,000 Words* (5) to determine how frequently each word appears in everyday usage. A number that indicated how often each word appears among a million written words was obtained. The seven most infrequently appearing words were then selected for each protocol and the frequencies of these seven words were totalled. Sixteen sums were thus obtained, one for each protocol. Rank-order estimates were made using the hypothesis that there would be an inverse relationship between the size of the sum and IQ.

The criterion measures of intelligence were based on full scale IQs on the Wechsler-Bellevue scale. Four Rorschach records were obtained from patients in each of the following IQ ranges: 120 and above, 110–119, 90–109, and 89 and below. Thus IQs were roughly rectangularly distributed, permitting comparisons of the accuracy of judgments based on the total protocols (Group III) with the results reported by Armitage *et al.* (1), who used a similar distribution. The protocols of male veterans who were referred for psychological evaluation during hospitalization were selected from the files at San Francisco Veterans Administration Hospital. The tests had been administered by eight different examiners. Patients suspected of having organic brain pathology were excluded from the sample. Only those cases whose Rorschach protocols were scored by the Klopfer and Kelley system (3) were used. The age of the patients in the sample ranged from 19 to 67, with a mean of 38.4. The mean number of years of education was approximately ten. From a diagnostic standpoint, the group was heterogeneous and included psychoneurotic, psychotic, personality disorder, and psychosomatic cases.

Table 1

Rho Correlations Between the IQ Estimates of Judges and Wechsler-Bellevue IQ

Judge	Group I	Group II	Group III
1	.79**	.85**	.79**
2	.76**	.81**	.67**
3	.74**	.66**	.64**
4	.68**	.63**	.61*

* Significant at or below the .05 level.
** Significant at or below the .01 level.

The 12 clinicians used as judges all had extensive training and experience in the use of the Rorschach and Wechsler-Bellevue. Most of them were graduate students nearing the completion of their Ph.D.s in clinical psychology.

Results

Correlation coefficients were used as indicators of the accuracy of judgments. In view of the rectangularity of the distribution, rank-order coefficients were employed. Table 1 indicates the correlations and the corresponding significance levels which were obtained by the judges. Results for Groups I and II indicate that estimates of intelligence can be made from Rorschach vocabulary. Since the accuracy of judges in Groups I and II is approximately the same, it does not appear to make any difference whether judges making their estimates from word lists have access to the protocols or not. It is to be noted that the judgments based on total protocol are no more accurate than those based solely on the vocabulary. In addition, the estimates based on the Thorndike and Lorge word frequency count resulted in a rho of .77 ($P < .01$), which is roughly comparable to the accuracy achieved by the judges. The results for Group III are in agreement with those of Armitage *et al.* (1), who report a median correlation of .69 for estimates based on evaluation of the total protocols.

Discussion

In comparison to the conventional Rorschach scoring variables whose contributions toward estimating intelligence have been quite limited, vocabulary is a variable that improves accuracy of judgments considerably. To rely solely on vocabulary in estimating IQ would simplify the task and might well be the most practical approach at present. The results suggest that the clinician is not indispensable for making the estimates. The Rorschachs could be submitted to a clerk who is instructed to select the rarest words from each protocol and to look them up in a word frequency table. After finding the seven most uncommon words in each, he totals their frequencies and enters this total into an empirically derived equation or table, which yields a numerical estimate of IQ.

The present investigation points the way toward areas of possible further research:

1. It is quite likely that vocabulary could also be evaluated from samples of verbal behavior other than the Rorschach. Thus estimates of intelligence might be made from the Thematic Apperception Test or from interviews, possibilities first suggested by Armitage *et al.* (1).

2. Since the Thorndike and Lorge word list is somewhat outdated and contains only 30,000 words, it is to be expected that if a more current and extensive word list could be utilized, it would lead to somewhat more accurate estimates of IQ than those reported here.

3. Since judges in Group III were not limited to specific cues in making their estimates of IQ, it is impossible to determine to what extent their accuracy of judgment was due to reliance on vocabulary. It would be informative to determine whether comparable accuracy would be achieved when vocabulary is eliminated as a cue. This could be accomplished by translating the protocols into basic English before giving them to the judges.

Summary

Three groups, each consisting of four judges, made estimates of Wechsler-Bellevue IQ from 16 Rorschach protocols. One group used word lists which they constructed from the protocols, another used word lists derived by others from the protocols, and the third used the protocols without instructions to focus on vocabulary. In addition, estimates were made in a more objective fashion using the Thorndike-Lorge word frequency count. The three groups and the objective approach achieved fairly comparable accuracy. Results suggest that IQ can be estimated as well, using only vocabulary, as by any other approach currently in use.

Received August 15, 1957.

References

1. Armitage, S. G., Greenberg, P. D., Pearl, D., & Daston, P. G. Predicting intelligence from the Rorschach. *J. consult. Psychol.*, 1955, **19**, 321–329.
2. Beck, S. J. *Rorschach's test.* Vol. II. *A variety of personality pictures.* New York: Grune & Stratton, 1949.
3. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book, 1942.
4. Terman, L. M., & Merrill, M. A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.
5. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia Univer., 1944.
6. Wechsler, D. *The measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Extraversion, Neuroticism, Radicalism, and Tendermindedness[1]

A. W. Bendig
University of Pittsburgh

Eysenck (2) has stated that the personality trait of Extraversion is negatively related to the social attitude dimension of Tendermindedness, but is not related to the attitude trait of Radicalism. Neuroticism is not related to either attitude dimension. However, Cattell states that Extraversion is negatively related to Radicalism (1, pp. 317–318) and also implies a relationship between Neuroticism and Radicalism (1, p. 209), but the direction of the correlation is left unclear. The present research was designed to clarify some of the interrelationships among these four variables.

The Maudsley Personality Inventory, containing scales of Extraversion and Neuroticism, was administered to 208 undergraduate *Ss* (117 men and 91 women) in five sections of introductory psychology. The scales of Radicalism and Tendermindedness developed by Melvin (2, pp. 276–280) were also administered to the same *Ss*. Comparisons of the mean raw scores of the two sex groups showed the female *Ss* to be more "tenderminded" than the male *Ss*, but the differences on the other three scales were not statistically significant.

Product-moment correlation coefficients among the four scales were separately computed within each sex group and the pairs of correlations tested for significant sex differences by the *r*-to-*z* method. The correlations were averaged when these tests did not approach significance. The average correlation between Extraversion and Neuroticism ($r = -.12$) was significant at the .10 level, but the correlations of Extraversion with Radicalism and Tendermindedness were not significant ($r = -.04$ and $-.01$). The average correlation between Neuroticism and Radicalism ($r = .22$) was significant at the .01 level and similar in both sex groups ($r = .18$ and .26). The correlations of Tendermindedness with Neuroticism and Radicalism were significant for the male *Ss* ($r = -.28$ and $-.19$), but not for the female *Ss* ($r = -.03$ and .08), and the sex differences in these correlations were significant at the .10 level ($z = 1.79$ and 1.95).

These results indicate that Eysenck's linking of Extraversion with Tendermindedness and Cattell's association of Extraversion and Radicalism were not confirmed by the data. However, Cattell's hypothesized relation of Neuroticism and Radicalism did appear within both sex groups of *Ss*. The sex differences among the correlations of Tendermindedness with Radicalism and Neuroticism may represent either a sampling fluctuation or a different organization of these three traits in men and women *Ss*.

Brief Report.
Received March 7, 1958.

[1] An extended report of this study may be obtained without charge from A. W. Bendig, Dept. of Psychology, University of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. Order Document No. 5578, remitting $1.25 for microfilm or $1.25 for photocopies.

References

1. Cattell, R. B. *Personality and motivation structure and measurement.* Yonkers, N. Y.: World Book, 1957.
2. Eysenck, H. J. *The psychology of politics.* London: Routledge & Kegan Paul, 1954.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Handwriting Conventions as Determinants of Human Figure Drawings[1]

Wayne Dennis
Brooklyn College

Projective theorists have suggested interpretations for many aspects of human figure drawings, among them the position of the drawing upon the page. Thus, Karen Machover (2) proposed that the location of the drawing toward the right side of the page indicates that the subject is environment-oriented, while a left-side location is characteristic of a self-oriented person. A drawing high on the page is related to optimism, a low-positioned drawing suggests depression. Levy (1) proposed that drawings on the left of the page indicate self-consciousness or introversion, toward the right, rebelliousness. For Levy, the use of the upper half of the page signifies striving toward high standards of achievement whereas persons who place drawings toward the lower edge of the sheet are likely to be stable, firmly rooted, or, in some cases, depressed. That these interpretations are based upon resemblances in meaning between words used to describe space (high, low, etc.) and words used to describe personality (elated, depressed, etc.) seems obvious. Whether such linguistic resemblances provide a valid basis for hypothesizing relationships between drawings and personality has not as yet been adequately determined.

[1] We wish to express our appreciation to the Lebanese, Egyptian, and American schools who provided data for this study. The Egyptian and Lebanese data were gathered in 1956–1957 when the author was a visiting professor at the American University of Beirut, and the gathering and analysis of these data were supported by a grant from the Rockefeller Brothers Fund. Special thanks are due to Habib Kurani and E. Terry Prothro of the American University of Beirut for their cooperation and encouragement and to Evelyn Raskin of Brooklyn College for consultation and suggestions.

Whatever may be the final decision in regard to the relationships between personality and graphic expression, it seems likely that not all of the variance present in drawings will be accounted for by individual differences in temperament and personality. There may be determinants of drawing position which are cultural and which have little if any relationship to personality. Culture determines the position on a page at which a person begins to write. If there is a habit-transfer between writing and drawing, the position of drawings will be affected by handwriting habits. This study examines the drawing location of two classes of writers; namely, those who begin to write at the top left of the page and those who begin to write at the top right of the page.

Method

Subjects

The writers who start at the top left of the page are represented by two subgroups, those writing English and those writing Armenian. In nationality, these *Ss* are Americans and Lebanese, respectively. Those starting at the top right of the page are Arabic writers. Their nationalities are Lebanese and Egyptian. All are children in elementary schools between ages five and ten.

The American *Ss* come from three widely separated localities: (*a*) a small town in Maryland, (*b*) a small town in Kansas, and (*c*) a suburban town near New York City. The Armenian *Ss* attended two schools in Beirut, Lebanon. The *Ss* comprised either complete samples or random samples from their respective schools. The total number of *Ss* in the above groups was 1,075.

Some of the Arabic-writing *Ss* were attending a public school in Port Said, Egypt. All others were attending public or private schools in Lebanon. These two groups totalled 651 *Ss*. These *Ss*, like the Armenian and American groups were, so far as could be determined, representative of their schools.

Testing and Scoring Procedures

The drawing test employed was the Goodenough draw-a-man test. Goodenough's instructions were followed. All drawings were made in pencil on 8½" by 11" white paper.

In scoring each drawing for the purposes of the present study, it was indicated, first, whether the topmost part was nearer or the bottom-most part was nearer their respective edges of the paper. Second, it was noted whether the left or right side of the drawing (left and right being designated from the observer's point of view) was nearer the lateral edges of the page. In judging the vertical location, the most extreme part of the drawing, such as the top of the hat or a projecting feather, was used. However, right and left judgments were based upon the most lateral parts of the *body*, excluding the extremities. In those drawings which lacked a body, judgments were based upon the head. When there was any doubt as to the judgments, measurements were made. When two measurements were identical, which seldom occurred, scores were assigned alternately to the two categories.

Results

In terms of the scoring method just described, a drawing must receive one of four scores: top left, top right, bottom left and bottom right. Top left means that the drawing was nearer the top of the paper than to the bottom, and nearer the left edge than the right edge. The other terms have corresponding meanings.

Table 1 shows the percentage of drawings falling into each category for each of the groups.

Table 1 shows that in the Armenian-American group, 48% of the drawings have a top-left position, whereas in the Arab group, 48% fall in the top-right group.

Table 1

Position of Figure Drawings on a Page

Location	Armenian-American	Arab
Top left	48%	24%
Top right	16%	48%
Bottom left	27%	14%
Bottom right	9%	14%
N	1,075	651

Discussion

Obviously, these findings are in accord with the habit-transfer hypothesis suggested earlier. Also in accord with this interpretation is the finding that 64% of the Armenian and American drawings and 72% of the Arab drawings are nearer the top of the page than they are to the bottom.

However, while *all* Armenians and Americans begin to write in the top-left portion of the page and *all* Arabic-writing *Ss* start their handwriting in the top-right quadrant, not all members of these groups place their drawings in the corresponding positions. Handwriting customs appear to have some influence in producing differences between certain groups, but they cannot, of course, explain the within-group variability. While the data here presented are consonant with the handwriting hypothesis, they are not entirely conclusive. The two groups compared no doubt differ in other respects as well as in regard to handwriting. It is unlikely that one can find two groups differing in handwriting but identical in all other respects. However, if one were to find that many groups differing in handwriting also differ in a consistent way in location of drawings, our hypothesis would seem to be firmly supported. Until a wider sampling of groups is obtained, the data here presented should serve as a caution in regard to interpreting cross-cultural differences in drawing locations as indicative of differences in personality. Such differences may be the result

of habit patterns having little relationship to emotional or social traits.

Summary

Evidence has been presented which is consonant with the hypothesis that the location of drawings depends in part upon whether subjects in writing their primary language begin at the top left or top right of the page. Further comparisons of other language groups are needed to test the hypothesis, but, meanwhile, caution in the interpretation of cross-cultural differences in location of drawings is indicated.

Received July 23, 1957.

References

1. Levy, S. Figure drawing as a projective test. In L. E. Abt & L. Bellack (Eds.), *Projective psychology*. New York: Knopf, 1950. Pp. 257–297.
2. Machover, K. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1948.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Relationship Between Edwards' and McClelland's Measures of Achievement Motivation[1]

Levon H. Melikian
American University of Beirut

Independent measures of achievement motivation have recently been developed by Edwards (1) and McClelland *et al.* (2). Edwards' measure of this variable is one of fifteen that have their origin in, and bear the same names as, a list of manifest needs presented by Murray (3). They form the Edwards Personal Preference Schedule which measures a "number of relatively independent normal personality variables" (1, p. 1). McClelland's method of measuring n Achievement is also rooted in Murray, since he uses imaginative stories evoked by two TAT cards and two other pictures designed by him and his associates. He has developed a quantitative method of scoring the imaginative material in terms of n Achievement categories (2, pp. 107–138).

Since both Edwards and McClelland independently measure the same variable, a high correlation is expected between the scores obtained on their two measures. The purpose of this study is to test whether such a relation exists or not.

Method

The subjects (*Ss*) in this experiment were 84 Arabic-speaking students in two psychology classes at the American University of Beirut. Approximately half of the *Ss* were citizens of Lebanon, and the other half came from Jordan, Syria, Iraq, and Bahrain. All the *Ss* had a good command of English, which is the language of instruction in the University. Sixty of the *Ss* were males.

[1] The author wishes to acknowledge his indebtedness to Aimee Sabbagh and Sumaya Kurani, of the Department of Psychology, for scoring the imaginative stories on the McClelland Test.

The full PPS was administered to the *Ss* as a class demonstration in the use of the questionnaire method. The *Ss* were told that *E* was interested in collecting norms on university students for help in his counseling work and were also told that these norms would be available to the class members. The *Ss* were told to use a private identification mark on their answer sheets and were asked to keep a record of it. A week later the four pictures used by McClelland were projected on a screen, and each *S* was given four sheets of paper, on each of which were printed the four standard questions used by McClelland about the pictures which the *Ss* were asked to answer. The instructions and method used by McClelland (2, p. 98) were followed throughout. The *Ss* were asked to use the same identification mark they had used on the PPS the week before. This was also presented to the class as a demonstration of projective tests which were to be discussed during the following class session. Both these tests were given under similar conditions, and the situation approximates McClelland's "neutral" condition of testing n Achievement.

The two tests were scored according to the methods prescribed by their authors. Fifteen scores were available for each subject on the PPS. The imaginative stories of the 84 *Ss* were independently scored by two research assistants and the interscorer reliability correlation was .93, which is considered to be satisfactory. Only the 69 *Ss*, 50 males and 19 females, who completed both tests were included in the final analysis. The age range for the 68 *Ss* who had indicated their age (on their last birthday) was 18 to 29, with a mean of 21.1 and a standard deviation of 2.2.

Results and Discussion

Table 1 shows the means and standard deviations on Achievement for our *Ss* and those of Edwards. The difference between the two means is not significant, *t* being .52, which points to the probable validity of this questionnaire for a non-American culture. To verify this, the Achievement items on the PPS were presented to three Arab professors in the fields of psychology, Arabic history, and Arabic literature. They were asked to judge whether any of the items were in any way alien to the Arabic culture and whether they could be taken as indices of achievement for the Arab group. There was full agreement among the judges about the validity of these items for an Arabic-speaking culture.

The mean score of the *Ss* on the imaginative stories was 2.22 with a standard deviation of 4.26. No comparable data are available for American groups. Even though such information is not available, a glance at the distribution of the different categories of achievement scores in Table 2 points to the possibility of cultural differences on certain categories. These differences will be discussed in a separate paper as they are not directly related to the present study. Evidence regarding the cross-cultural validity of the scoring method used is presented by McClelland (2, pp. 168–172).

The correlation between the scores of the 69 *Ss* on the PPS achievement variable and the McClelland achievement scores is .16. Since an *r* of .16 is not significant, it indicates an absence of a relationship between two methods that claim to measure the same variable.

Like all motives, the achievement motive is rather complex. McClelland (2, p. 217) describes his score as that of a "composite index of a variety of types of achievement motivation" which he demonstrates as having a meaningful relationship to varied types of behavior. Among the theoretical distinctions which he makes between different types of achievement is one type that is characterized chiefly by a "hope of success" and another by a "fear of failure" (2, p. 215).

This theoretical distinction may possibly account for the absence of a relationship between the two measures. An inspection of the PPS items which measure achievement motivation shows that they refer to competition with some standard of excellence similar to the "hope for success" type of achievement. "To do one's best, to be successful, to accomplish . . . to do . . . , to solve . . . , to do things better than others . . . etc. . . ." are given by Edwards (1, p. 5) in describing his Achievement variable.

Besides the above, responses to the Edwards items are made to clearly structured stimuli, whereas the imaginative stories are responses to a variety of cues both internal and external, in the environment and the pictures, as well as to experimentally induced cues. This difference points to the probability that the two measures tap different levels of

Table 1

Mean Achievement Scores of Arab and American[a] Subjects on the PPS

Group	*N*	Mean	*SD*
Arabs	84	14.65	4.26
Americans	1509	14.38	4.36

[a] U. S. scores taken from Edwards (1, p. 6).

Table 2

Frequencies of Various Achievement Categories[a] under Neutral Experimental Conditions for Both Arab and American Subjects[b]

Achievement category	Americans	Arabs
Achievement Imagery (AI)	75	70
Doubtful Imagery (TI)	49	102
Unrelated Imagery (UI)	49	112
Need (N)	35	25
Instrumental Activity (I)	31	33
Positive Anticipatory Goal State (Ga +)	25	14
Negative Anticipatory Goal State (Ga −)	20	12
Personal Obstacle (Bp)	8	8
Environment Obstacle (Bw)	17	10
Positive Affect State (G +)	17	16
Negative Affect State (G −)	23	13
Nurturant (Nup)	5	5
Achievement Theme (Ach. Th.)	62	58

[a] The total number of stories for the Arab group is 276 (69 *Ss*, 4 stories each); and for the American group is 156 (39 *Ss*, 4 stories each).
[b] Data on the American *S* is taken from McClelland (2, p. 141).

achievement, a manifest level, and a latent level (2, pp. 196–217), and may account for the absence of a relationship between the two measures.

In the light of the findings, and in the interest of avoiding confusion, it is suggested that the two measures of achievement motivation be either renamed or qualified.

Summary

The PPS and the McClelland pictures for measuring achievement were given a week apart to a group of 84 Arabic-speaking *Ss* at the American University of Beirut. No relationship was found between the achievement scores of the 69 *Ss* who completed both tests. The results were discussed in terms of the complexity of the achievement motive and, also, of the probability that the two methods measure the achievement motive at different levels.

Received August 2, 1957.

References

1. Edwards, A. L. *Manual: Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.
2. McClelland, D., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
3. Murray, H. A., *et al. Exploration in personality.* New York: Oxford Univer. Press, 1938.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Second-Order Factors in Personality Measurement

Samuel Karson
University of New Hampshire

and Kenneth B. Pool[1, 2]
University of California, Berkeley

It is surprising that to date so few published studies on the Sixteen Personality Factor Questionnaire have appeared which have not emanated from the Laboratory of Personality Assessment at the University of Illinois. The reasons for this lack of popularity are not readily apparent, since Cattell and his coworkers have published widely and frequently (3, 4, 5, 7, 8). Karson and Pool (11) found much similarity between the 16 P.F. and the MMPI which they interpreted as offering support for the construct validity of the 16 P.F. test. Their major criticism of the test was that the coefficients of equivalence of the factors on Forms A and B were not nearly high enough to justify their being used as equivalent forms, and that even after being corrected for full length with the Spearman-Brown prophecy formula, the resultant coefficients of half of the 16 factors were still not high enough to warrant their use for accurate individual personality description.

In a recent paper, Cattell (5) reported four second-order personality factors which he identified as Anxiety-vs.-Dynamic Integration, Extraversion-vs.-Introversion, Cyclothyme-vs.-Schizothyme Constitution, and Unbroken Success-vs.-Frustration. These findings were based on a sample of 408 normal subjects composed of 181 undergraduate students of both sexes and 227 U.S. Air Force young men in training. The aim of the present study was to attempt to replicate the second-order factors described by Cattell in order to determine the degree of stability of these factors over different samples. From a theoretical viewpoint, if different samples resulted in yielding unrelated factor structures, then the argument (4, p. 3) that factors actually correspond "to psychological functional unities rather than to mere mathematical conveniences" would be highly questionable.

Sample

The sample consisted of 71 USAF officers who were referred to the School of Aviation Medicine for medical and psychological evaluation. The average age was 33, with a range of 20 to 39 years, and the mean number of years of education was 14, with a range of 12 to 16 years. The mean Full Scale IQ on the Wechsler Adult Intelligence Scale was 121.98, with a standard deviation of 6.50. The patients were referred for evaluation between August, 1955 and July, 1956. The sample employed in this research is believed to be typical of the flying personnel who are referred for psychiatric consultation in the U.S. Air Force.

None of the patients was psychotic, but almost all were manifesting some disturbance in functioning efficiency which was impairing, at least temporarily, their ability to perform their duties as USAF flying personnel. Upon clinical examination, which included psychiatric and psychological evaluation, there was general agreement that anxiety and its disrupting effects usually played a key role in the etiology of the patients' symptomatology,

[1] The data for this study were collected and the statistics computed while both authors were at the School of Aviation Medicine, USAF.

[2] A summarized version of this paper was presented at the 1957 meetings of the APA in New York City by the senior author.

Table 1

Intercorrelations* of Factor Scores on the Sixteen P.F. Test Forms A and B Combined ($N = 71$)

Score	*A*	*B*	*C*	*E*	*F*	*G*	*H*	*I*	*L*	*M*	*N*	*O*	Q_1	Q_2	Q_3	Q_4
A		.02	.12	.13	.42	−.10	.45	−.12	−.07	−.08	.12	−.07	.01	−.27	−.08	.11
B			−.12	.08	.00	−.08	−.11	.02	.21	.18	−.13	.21	−.15	.02	−.15	.20
C				.22	.30	−.17	.41	−.38	−.59	−.57	.03	−.71	−.40	−.13	.43	−.64
E					.41	−.03	.49	−.43	.00	−.07	.13	−.32	.12	−.01	−.13	−.08
F						−.20	.66	−.23	−.14	−.14	.12	−.28	−.02	−.16	−.19	−.13
G							.08	.16	.23	.03	−.07	.00	.07	.26	.17	.04
H								−.26	−.23	−.27	.18	−.52	−.17	−.36	.05	−.28
I									.20	.45	−.23	.40	.04	.16	−.09	.24
L										.52	.00	.46	.22	.22	−.18	.51
M											−.13	.56	.31	.29	−.36	.53
N												−.18	.11	−.01	−.03	−.12
O													.21	.23	−.41	.76
Q_1														.24	−.16	.17
Q_2															.01	.09
Q_3																−.42
Q_4																

* For significance at the .05 level of confidence a Pearson r must equal .24.

even though an actual incapacitating, diagnosable neurosis was seldom present. It should be remembered that the sample was composed of mentally and physically superior experienced military flying officers, many of whom had successful combat records in either World War II or Korea.

Method and Results

The 16 P.F. test (Forms A and B) was administered as a part of the routine diagnostic psychological test battery. The two forms of the test were alternated so that about one-half of the sample took Form A first, while the other half had Form B initially. Both forms of the test were usually administered on the same day, one in the morning and the other in the afternoon. The tests were scored according to the directions given in the handbook. The raw scores were converted into standard scores by using the sample of male college students and high school seniors whose average age was 21 as the reference group. Pearson coefficients of correlation among the standard scores for the 16 factors were computed to achieve the matrix of intercorrelations presented in Table 1.[3]

The factor analysis was then done, using Thurstone's centroid method (12). As guides for when to stop factoring, criteria suggested by Burt and Banks (1) and Guilford and Lacey (10) were used. After six factors had been extracted, an estimated 109% of the common factor variance had been accounted for, and the sixth factor residuals ranged in size from − .068 to .093. Rotation to simple structure was then attempted without regard to positive manifold. As a guide in rotation, the goal was to maximize the number of variables with zero projections on the reference vectors. The factors were rotated blindly two at a time by means of graphs and six independent, second-order factors were extracted.[4] The new factor loadings and their communali-

[3] Although this procedure for obtaining the matrix of intercorrelations is admittedly not as precise as when one begins with the cosine matrix among the simple structure primary factors and then inverts this matrix to obtain the correlations among the factors, Cattell has previously employed this method (4, p. 18). On this same issue, Cattell (5, p. 412) states: "However, it must be granted that where factors are already very purely measured by existing tests, with no systematic bias and with effective suppression of peer factors in each factor score, the factoring of actual factor scores, i.e., of one variable for each factor, may yield a reasonable approximation to second-order factors. This seems to have happened with the 16 P.F. test in the large Air Force study."

[4] The centroid factor matrix, the matrix of sixth factor residuals and the direction cosines have been deposited with the American Documentation Institute. Order Document No. 5586, remitting $1.25 for microfilm or $1.25 for photocopies.

Table 2

Rotated Second-Order Factor Loadings

First order factor	I	II	III	IV	V	VI	h^2
A	.03	.51	.09	.34	.10	−.02	.3951
B	.24	.03	.38	−.10	.03	−.07	.2187
C	−.84	.24	.27	−.17	.05	−.08	.8739
E	−.04	.54	−.02	−.33	.04	−.37	.5410
F	−.11	.75	.04	−.02	.20	−.03	.6175
G	.03	−.06	−.08	.01	−.61	.08	.3895
H	−.33	.84	−.04	.19	−.16	.00	.8778
I	.32	−.23	−.01	.03	−.09	.60	.5244
L	.65	−.02	.00	−.08	−.41	−.08	.6038
M	.72	.02	.00	−.23	−.04	.32	.6757
N	−.07	.11	−.30	.07	.01	−.27	.1849
O	.81	−.34	.18	.12	.09	.09	.8347
Q_1	.34	.02	−.47	−.17	−.02	−.05	.3687
Q_2	.21	−.23	−.05	−.47	−.26	−.01	.3881
Q_3	−.54	−.19	.03	.01	−.34	.04	.4459
Q_4	.80	−.06	.21	.19	−.04	−.04	.7270

ties after rotation of axes are presented in Table 2.

Description of the Factors

The chief indications of the nature of Factors 1 through 6 are presented below with a contingent title given to each factor.

Factor 1. Anxiety-vs.-Dynamic Integration

Symbol	*16 P.F. Handbook Title*	*Loading*
C	Ego Strength	−.84
O	Anxious Insecurity	.81
Q_4	Nervous Tension	.80
M	Hysteric Unconcern	.72
L	Paranoid Schizothymia	.65
Q_3	Will Control & Character Stability	−.54
Q_1	Radicalism	.34
H	Adventurous Autonomic Resilience	−.33
I	Emotional Sensitivity	.32

This factor shows a marked resemblance to the second-order factor reported by Cattell (5) on which Q_3, *O*, Q_4, *L*, *C*, and *H* loaded heavily. He identified this factor as Anxiety-vs.-Dynamic Integration and we have adopted his title. Further support for this identification is forthcoming from an earlier study by Cattell (4) in which a factor described by him as Anxiety-vs.-Self-Assuredness was reported. The 16 P.F. test factors, Q_4, *O*, Q_3, *C*, *L*, *M*, and *H* all loaded on this factor.

Additional support is provided by Karson and Pool (11) who reported that the MMPI psychasthenia (*Pt*) scale correlated significantly with factors *C*, *O*, Q_4, *M*, *L*, Q_3, *H*, and *I*. Moreover, the factors with significant loadings on Factor 1 correlated more highly with the *Pt* scale than any of the other 13 MMPI scales, with the possible exception of *Sc* which correlated .89 with *Pt* for the present sample (11). High correlations between *Pt* and the Taylor Anxiety scale have been reported by Eriksen and Davids (9) and by Brackbill and Little (2).

Factor 2. Extraversion-vs.-Introversion

Symbol	*16 P.F. Handbook Title*	*Loading*
H	Adventurous Autonomic Resilience	.84
F	Surgency	.75
E	Dominance	.54
A	Cyclothymia	.51
O	Anxious Insecurity	−.34

Factor 2 also shows much similarity to another of Cattell's second-order factors on which *M*, *F*, *A*, Q_1, Q_2, and *H* had significant loadings. He identified this factor as Extraversion-vs.-Introversion, and once again we have adopted his title (5). Cattell reports that in still another study (5, p. 416) the *E* factor (which we found) also enters in and, moreover, that an alternative rotation which was considered for his Factor 2 on 4 would also have done so again. The reasons why Q_1 did not load on Factor 2 in the present study, while *O* did, are unknown, although it is believed that sampling differences may be responsible for these variations.

Karson and Pool (11) provide additional support for the identification of Factor 2 inasmuch as they reported significant correlations between certain 16 P.F. scores and particular MMPI scale scores which appear relevant to this identification.

The present findings do not, however, support Cattell's assertion (5, p. 416) that the "relatively low loadings of the more hereditarily determined *H* and *A* factors suggest that Factor 2 is an environmentally determined pattern." Factor *H* had the highest loading of all in the present study. Nevertheless, his conclusion (5, p. 417) that "There now seems ground for recognizing extraversion-introversion as something more than a correlation cluster or surface trait, and perhaps for specializing the term for this measur-

able second-order factor" is in agreement with the present findings.

Factor 3. Unbroken Success-vs.-Frustration

Symbol	*16 P.F. Handbook Title*	*Loading*
Q_1	Conservatism	−.47
B	General Intelligence	.38
N	Lack of Sophistication (Naivete)	−.30
C	Ego Strength	.27

Factor 3 shows some resemblance to Cattell's factor of Unbroken Success-vs.-Frustration inasmuch as Q_1 and *N* are present in both patterns as well as high intelligence from an objective behavior study (5, p. 417).

Karson and Pool (11) demonstrated, however, that for the present USAF officer sample, Factor B was not a valid measure of intelligence. They also reported significant correlations between *B* and *Mf* and *L*, respectively, and between Q_1 and *F*, *Hy* and *Mf* from the MMPI. This pattern of relationships makes Factor 3 difficult to identify. The present authors have chosen to adopt Cattell's title of Unbroken Success-vs.-Frustration, although a rival hypothesis of Spontaneity-vs.-Inhibition was also entertained.

Factor 4. Independence-vs.-Dependence

Symbol	*16 P.F. Handbook Title*	*Loading*
Q_3	Lack of Resolution	−.47
A	Cyclothymia	.34
E	Lack of Dominance (Submission)	−.33

Factor 4 has some similarity to a second-order factor (4, p. 40) of Schizothyme Tenacity-vs.-Cyclothymia on which Factors *E*, *F*, *H*, *A*, and Q_2 loaded. This identification is difficult to reconcile, however, with a more recent second-order factor again identified by Cattell (5, p. 417) as Cyclothyme-vs.-Schizothyme Constitution on which Factors *A*, *I*, *N*, Q_3, and *C* loaded significantly.

Karson and Pool (11) reported significant correlations between Q_2 and *Si* and *Mf*, between *A* and *Si*, and between *E* and *Mf*, *Si*, and *Sc*, respectively. In view of this pattern of interrelationships, Factor 4 has been tentatively identified as Independence-vs.-Dependence. The handbook (8) for the 16 P.F. lends support for this hypothesis by its description of the Q_2 − person (8, p. 11), the *A* − and *A* + person (8, p. 6), and Factor *E* as well (8, p. 7).

Factor 5. Sociopathic Deviancy

Symbol	*16 P.F. Handbook Title*	*Loading*
G	Lack of Super-Ego Strength	−.61
L	Lack of Paranoid Schizothymia	−.41
Q_3	Lack of Will Control	−.34
Q_4	Lack of Resolution	−.26

The identification of this factor is especially difficult, principally because the two factors with the highest loadings, namely *G* and *L*, are among the ones about which least is known. Karson and Pool (11) reported that none of the MMPI scale scores correlated significantly with Factor *G* which, strangely enough, had its highest correlations with *Pd* and *K*. These findings as well as the authors' clinical experience with this test over a two-year period raise some doubt as to the identification of *G* as a factor of "effective" super-ego pattern (6). Factor *L* is also puzzling; it did not correlate with *Pa;* however, it correlated more highly with *Pt* and *Sc* than did *Pa*. Moreover, the high loading of *L* on the second-order factor of Anxiety-vs.-Dynamic Integration, which has been found consistently (4, 5, 7), does not appear to have been satisfactorily explained on a theoretical basis although such an explanation has been attempted (7, p. 230).

Two hypotheses are suggested for the identification of Factor 5; first, that it is a factor of sociopathic deviancy and, second, that it represents a lack of obsessive-compulsive personality characteristics in our sample. The latter interpretation would necessitate a reexamination of Factor *L* with emphasis given to its anxiety and rigidity aspects rather than to paranoid suspicion. At any rate, it is encouraging to see some relationships emerge between the highly esoteric world of factor analysis and the more familiar conceptualizations of psychiatric nosology.

Factor 6. Latent Homosexuality (Passive)

Symbol	*16 P.F. Handbook Title*	*Loading*
I	Emotional Sensitivity	.60
E	Lack of Dominance (Submission)	−.37
M	Hysteric Unconcern	.32
N	Lack of Sophistication (Naivete)	−.27

The particular constellation of loadings on this factor suggests the hypothesis of latent homosexuality of a passive nature. Some sup-

port for this identification comes from the significant correlation reported between *I* and *Mf* (11) and from the description of the *I* + person presented in the 16 P.F. handbook (8, p. 9). This factor may serve to throw light on the psychodynamics of military flyers, particularly with regard to the frequently made clinical observation of "masculine strivings" for this group.

Some mention should be made of the fact that Cattell (5) used an oblique solution in obtaining the four second-order factors which he described, while the present study reports six independent factors. These differences notwithstanding, the present findings appear to be encouraging for those who believe in the stability and reality of factorial solutions. Why two of Cattell's factors correlated .43, while the same two factors were not significantly correlated in the present study, is not readily apparent although this discrepancy may be a function of methodological differences, sampling differences, or both.

Summary

In order to evaluate the stability of factorial solutions over different samples, an attempt was made to replicate the second-order personality factors reported by Cattell. Forms A and B of the 16 P.F. were administered to a sample of 71 USAF officers on flying status, who had been referred to the School of Aviation Medicine, USAF, for medical and psychological evaluation. The test scores were punched on IBM cards, and a correlation matrix was derived using Pearson product-moment correlation coefficients. The factor analysis was done using Thurstone's centroid method. Rotation to simple structure by means of graphs yielded six orthogonal, second-order personality factors. These were described as follows: 1. Anxiety-vs.-Dynamic Integration; 2. Extraversion-vs.-Introversion; 3. Unbroken Success-vs.-Frustration; 4. Independence-vs.-Dependence; 5. Sociopathic Deviancy; 6. Latent Homosexuality (Passive). The first three factors named above appear to be replications of three of the four second-order factors reported by Cattell and his titles for them were used. The remaining differences in the nature and number of second-order factors found between the two studies can probably be explained at least partially by methodological differences, by the varying degree of simple structure actually attained, and by differences in age, status, and sex as well as in the personality characteristics of the two samples.

Received August 5, 1957.

References

1. Burt, C., & Banks, Charlotte. A factor analysis of body measurement for British adult males. *Ann. Eugen., Camb.*, 1947, **13**, 238–256.
2. Brackbill, G., & Little, K. B. MMPI correlates of the Taylor Scale of Manifest Anxiety. *J. consult. Psychol.*, 1954, **18**, 433–436.
3. Cattell, R. B. *Personality: A systematic theoretical and factual study*. New York: McGraw-Hill, 1950.
4. Cattell, R. B. Psychiatric screening of flying personnel. *Personality structure in objective tests —A study of 1,000 Air Force students in basic pilot training. USAF Sch. of Aviat. Med.*, Proj. No. 21-0202-0007, Feb., 1955, 1–50 (Rep. No. 9).
5. Cattell, R. B. Second-order personality factors in the questionnaire realm. *J. consult. Psychol.*, 1956, **20**, 411–421.
6. Cattell, R. B. *Handbook for the I.P.A.T. Anxiety Scale*. Champaign, Ill.: Institute for Personality and Ability Testing, 1957.
7. Cattell, R. B. The conceptual and test distinction of neuroticism and anxiety. *J. clin. Psychol.*, 1957, **13**, 221–233.
8. Cattell, R. B., Saunders, D. R., & Stice, G. F. *The Sixteen Personality Factor Questionnaire*. Champaign, Ill.: Institute for Personality and Ability Testing, 1949.
9. Eriksen, C. W., & Davids, A. The meaning and clinical validity of the Taylor Anxiety Scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.
10. Guilford, J. P., & Lacey, J. I. *Printed classification tests*. Army Air Forces Aviat. Psychol. Progr. Res. Rep. No. 5. Washington, D. C.: U. S. Gov't Printing Office, 1947.
11. Karson, S., & Pool, K. B. The construct validity of the Sixteen Personality Factors Test. *J. clin. Psychol.*, 1957, **13**, 245–252.
12. Thurstone, L. L. *Multiple factor analysis*. Chicago: Univer. Chicago Press, 1947.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The MMPI Performance of Veterans with Organic and Psychiatric Disabilities[1]

Joseph J. Motto
Detroit Regional Office, Veterans Administration

The purpose of this paper was to investigate the Minnesota Multiphasic Personality Inventory (MMPI) performance of two groups of disabled veterans to determine whether there were areas of significant difference in terms of personality functioning. Particular attention was devoted to the *K* correction scale to determine if significant differences existed in the "test-taking attitude" of veterans with psychiatric as compared to those of organic disabilities.

The setting of the study was a VA counseling service for seriously disabled veterans. The *Ss* consisted of a group of 50 veterans having VA ratings for tuberculosis and gunshot wounds and a like number of veterans with VA ratings for schizophrenia. The *Ss* were selected at random from the population of male counseling cases seen within the past three years.

The MMPI scores of the psychiatric (P) group were compared with the organic (O) group, and both of these groups were compared with the MMPI standardization group on several of the scales, and with a group of nondisabled "normal" veteran controls.

Examination and statistical analyses of the data revealed the following:

1. MMPI scales, *Pt* and *Sc*, significantly differentiated those of organic disabilities from those of schizophrenic disabilities.

2. In terms of personality characteristics measured by the MMPI, the organic group was more similar to the psychiatric group than they were to a group of nondisabled, "normal" veteran controls.

3. The presence of an organic disability resulted in *Hs* scores which were not significantly higher than those produced by the psychiatric group.

4. The *K* scale functioned as predicted by test authors and conformed to their rationale of the suppressor variable when applied to those with organic disabilities, the sole exception being the *Mf* scale where it appears a *K* correction weight is needed for this group. A similar trend to an inverse relationship between *K* and *Mf* was seen in the P group.

5. There were indications of the need for an additional *K* correction weight for veterans with psychiatric disabilities on *Sc*, *Ma*, and *Pa* scales.

6. Compared to the control group, both disability groups produced higher *F* scores. The organics differed from the normals by the presence of an elevated neurotic triad, whereas the psychiatric group differed from the normals by elevations of both the neurotic and the psychotic triads.

7. *K* was found to have a high positive relationship with *L* and a significant inverse relationship with *F*.

[1] An extended report of this study may be obtained without charge from Joseph J. Motto, 8635 Elmira, Detroit 4, Michigan, or for a fee from the American Documentation Institute. Order Document No. 5579, remitting $1.75 for microfilm or $2.50 for photocopies.

Brief Report.
Received March 10, 1958.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Two New Experimental Scales of the MMPI

Kenneth B. Little[1]
Stanford University

and Jerome Fisher
VA Hospital, San Francisco, California

Several investigators (1, 2, 3) have made use of one or both of two experimental scales derived by us in the course of a cluster analysis of the Hysteria scale of the MMPI. The purpose of this note is to describe briefly the development of these scales and to indicate some of their correlates in clinical and experimental work.

As part of the study of differential personality patterns in diagnostic center and general medical patients in a Veterans Administration hospital, it was observed that the Hysteria scale score was the highest point on the MMPI profiles of both groups. In both, the mean *T* score exceeded 70 despite the fact that the general medical patients were hospitalized for nonpsychiatric disorders. The observation of an elevated *Hy* score in general medical patients is, of course, a common one. On the assumption that these elevated scores had a different meaning in a nonpsychiatric population than in one with some history of psychologic disorder, we initiated a cluster analysis of the matrix of tetrachoric intercorrelations among the 60 Hysteria scale items. The patient population on which the correlations were derived consisted of 60 diagnostic center neuropsychiatric subjects and 90 random admissions to the general medical wards.

Nine relatively clear-cut clusters were obtained. Five of these clusters had moderate to high positive correlations among themselves and zero or slightly negative correlations with the remaining four. The latter group of clusters had, conversely, moderate to high positive correlations within itself. The findings suggested, of course, that two primary dimensions underlay the cluster structure. The group of five clusters (comprising 23 of the *Hy* items) was composed for the most part of statements which consist largely of physiologic symptoms to which the subjects "admitted." This group was labeled the Admission (*Ad*) scale.

The second scale was composed of the 19 items falling into the group of four clusters. These items consisted for the most part of items scored if the subject denied their validity for him. In content, they were statements about poor interpersonal relations, feelings of hostility, suspicion, and the like. The second cluster was therefore called the Denial (*Dn*) scale. As a final step, biserial correlations were computed between each of the remaining 18 Hysteria scale items and the *Ad* and *Dn* scale scores. Items were added to the scale with which they had the highest correlation. Two items correlated zero with both scales and were discarded.

The item composition of the two scales (as they are numbered in the group form booklet of the MMPI) is as follows: *Ad* (32 items)—Items 2, 3, 7, 8, 9, 55, 103, 107, 128, 137, 153, 160, 163, 174, 175, 188, 190, 192, 230, 243 scored if marked *False*, and items 10, 23, 32, 43, 44, 47, 76, 114, 179, 186, 189, and 238 scored if marked *True*. *Dn* (26 items)—Items 6, 12, 26, 30, 71, 89, 93, 109, 124, 129, 136, 141, 147, 162, 170, 172, 180, 201, 213, 234, 265, 267, 279, 289, 292 scored if marked *False*, and item 253 scored if marked *True*. All 58 items are scored in the same direction as they are scored on the *Hy* scale.

[1] Now with the Training and Standards Branch, National Institute of Mental Health, Bethesda, Maryland.

Correlations between the new scales and other MMPI scales for two different groups of subjects,[2] $N = 100$ in each group, are given in Table 1. The Kuder-Richardson reliability coefficients based on all 200 cases were *Ad* = .83 and *Dn* = .75. *Ad* has its highest correlation with the MMPI *Hs* scale, a not too surprising observation in view of the 18-item overlap between the two scales. In most populations, *Hs* and *Ad* could be substituted one for the other. The *Dn* scale has as its highest correlation that with the MMPI *K* scale, with which it shares nine items in common.

In the clinical application of these scales, it has been found that high scores on the *Dn* scale generally describe the uninsightful, the anti-intraceptive, the morally virtuous individual; high scores on *Dn* are also seen with "muted" or pseudo-normal profiles. Extreme scores on the *Ad* scale, on the other hand, depict the person in general psychologic distress who complains and is anxious about his somatic functions and who also reports disturbances in his object relationships. Such high scores on *Ad* are usually seen with overall elevated MMPI profiles. When both scales have moderate to high scores, equivalent to an elevation on the original *Hy* scale, there is the distinct clinical impression of the presence and operation of conversion neurotic dynamics in the behavior and history of the individual.

Table 1

Correlations of Admission and Denial Scales with MMPI Scales for Two Populations

MMPI scale	General medical patients		Patients referred to psychiatric service	
	Ad	*Dn*	*Ad*	*Dn*
L	−09	35	−10	41
F	52	−23	64	−54
K	−33	78	−57	88
Hs	89	−29	90	−42
D	59	−08	70	−41
Pd	43	−27	46	−37
Mf	17	01	17	−14
Pa	22	17	48	−23
Pt	57	−60	75	−71
Sc	53	−43	75	−65
Ma	19	−46	57	−59
Si	44	−49	55	−62
Ad	—	−30	—	−56
Dn	−30	—	−56	—

[2] The subjects were tested as part of a project supported by a grant (M-1132) from the National Institute of Mental Health.

Received July 29, 1957.

References

1. Block, J., & Thomas, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254–259.
2. Gordon, J. Interpersonal predictions of repressors and sensitizers. *J. Pers.*, in press.
3. Truax, C. B. The repression response to implied failure as a function of the hysteria-psychasthenia index. *J. abnorm. soc. Psychol.*, 1957, **55**, 188–193.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Parental Attitudes of Mothers of Schizophrenics[1]

Marvin Zuckerman
Institute of Psychiatric Research, Indiana University Medical Center, Indianapolis

Mary Oltean
Carter Memorial Hospital

and Irwin Monashkin
Purdue University

Psychoanalytic theorists such as Fromm-Reichman (3), Rosen (6), and Arieti (1) have hypothesized that schizophrenia has its origin in the early relationship between mother and child. They imply that the mother of a schizophrenic child has certain destructive types of attitudes toward children which she conveys to the child through her behavior and verbalizations. Unfortunately, we must rely on retrospective evidence bearing on this hypothesis. Tietze (9) interviewed a group of mothers of schizophrenics and characterized them as overanxious, obsessive and domineering. However, no control group was compared with these mothers, and one may question the objectivity of an open interview with a mother already identified as the mother of a schizophrenic.

The need for more objective techniques to assess parental attitudes encouraged other researchers to develop objective instruments. The typical parental attitude survey uses items which are stated in the form of generalities and clichés in order to make them sound socially acceptable. The parent is asked to state whether she mildly or strongly agrees or disagrees with these items. Freeman and Grayson (2) used the Shoben Parent-Child Attitude Survey in comparing 50 mothers of schizophrenics with 50 control mothers. They found the mothers of schizophrenics scored significantly higher on the Possessive and Ignoring scales (on the latter scale they accepted the .10 level of confidence) than the control mothers. Unfortunately, they did not control educational or socioeconomic differences between their groups. From the way in which the control group data was gathered, it is likely that differences existed in the educational levels of the groups. Furthermore, since they relied on volunteer mothers solicited by volunteer students, various selective factors could have determined the scores in the control group.

Mark (4), in a more carefully controlled study, found that a significant number of attitude items distinguished a group of 100 mothers of schizophrenics from 100 control mothers. However, McFarland (5) attempted to replicate this study with the same items and failed to find a significant number of items distinguishing the two groups.

The purpose of the present study was to retest the hypothesis that mothers of schizophrenics have more severe (controlling and rejecting) parental attitudes than mothers of normals.

Procedure

The Parental Attitude Research Instrument (PARI) was the main instrument used in this study. This test was developed by Schaefer and Bell (7), with items modeled after the types used by Mark (4) and Shoben (8). An attempt was made to develop a number of

[1] The authors are grateful to Donald Moore and Margaret Morgan for their permission in making the facilities of Carter Memorial Hospital available for this research. Harry Brittain assisted in the analysis of the data.

short, but internally consistent, scales measuring specific types of attitudes, to permit somewhat broader interpretations than can be made from *post hoc* item analyses. The test consists of 23 scales, each of which measures some attitude theoretically relevant to personality development in children. Normative data are furnished by Schaefer and Bell (7), and in a previous article by Zuckerman, Ribback, Monashkin, and Norton (11). A factor analysis of the scales using a heterogeneous sample, including the group reported in this paper, can be found in the latter article. The factors found in this study were used to derive summary scores in the current study by weighing scales roughly by their factor loadings. Factor A seems to be an Authoritarian-Controlling factor which loads heavily on 16 of the scales. Factor B seems to be a Hostility-Rejection factor which loads on three scales. Factor C seems to be a Democratic Attitude factor which loads on three scales.

Subjects

Forty-seven mothers of male and female patients diagnosed as schizophrenic in an acute treatment hospital were given the PARI. Ratings of pathology on 30 cases of mixed diagnoses were available from a previous study and provided some information on the reliability of diagnoses at the institution. A seven-point rating scale ranging from mildly neurotic to severely psychotic was used in rating the case-history summaries of the patients. Ratings were global, but factors considered included severity of symptoms, chronicity of the disorder, response to treatment, and premorbid adjustment. Two of the authors made the ratings independently, and their ratings correlated .80. The agreement between the raters and the diagnosis assigned the patient is summarized in Table 1. It can be seen that there is high agreement between psychiatric diagnosis and the ratings except in the middle of the scale where the borderline cases would be expected to fall. Although there is probably some contamination between the two sets of data, the experimenters felt justified in using the psychiatric diagnosis of schizophrenia as a criterion, recognizing that wide behavioral differences may exist within the disorder diagnosed as "schizophrenia."

Table 1

Relation Between Case History Rating and Psychiatric Diagnosis

Psychiatric diagnosis	Case history rating		
	1–3	3.5–4.5	5–7
Nonschizophrenic	8	2	1
Schizophrenic	0	2	17

Note.—Chi square, corrected (combining cells) = 15.56. $df = 1$, $p < .001$.

Table 2

Composition of Mothers of Schizophrenic and Mothers of Normal Groups

Level of education	Schizophrenics' mothers		Normals' mothers	
	N	Mean age	*N*	Mean age
Grammar school	11	52.4	11	52.0
High school	24	48.0	24	48.8
College	7	50.7	7	49.1
All cases	42	49.6	42	49.7

A group of 47 mothers of normals was derived from the older age range of the normal sample (11) and from some additional normals gathered at a later date. The control group consisted of mothers given the PARI at church social and nursery school meetings, with some mothers of students from a night extension college.

Since education of the mother seems to influence most of the scales and one of the factors (11), it was put into the analysis of variance design using three levels: grammar school, high school, and college. Age differences were then compared between diagnostic groups across levels. At the grammar school level, there was a significant age difference; therefore, the five oldest mothers of schizophrenics and the five youngest mothers of normals were discarded. The resultant samples, containing 42 mothers in each group, are described in Table 2. Sex of the patient was considered before grouping mothers of male and female schizophrenics together. Twenty-one mothers of male schizophrenics

were compared with 21 mothers of female schizophrenics. A significant difference was found on only one scale. Therefore, the groups were combined. Other variables considered were the child's diagnostic subclassification, paranoid vs. other schizophrenics, and the time interval between the admission of the patient and the administration of the test to the mother. Neither of these variables seemed to markedly influence the average parental attitude scores.

Results and Discussion

The results of the analyses of variance on Factors A, B, and C, and on 20 of the individual scales are reported in Table 3. Only one of the differences between groups was significant. The mothers of schizophrenics scored significantly lower on Scale 8, strictness, than the controls. However, 1 out of 23 comparisons may be considered a finding attributable to chance. The differences between education levels were significant for Factor A and 14 of the 20 individual scales. The interaction effects were significant for Factor A and eight of the individual scales. An analysis of the interaction effect in Factor A can be seen in Table 4. At the grammar school level, the mothers of schizophrenics tend to score higher than the controls, and at the high school level they scored significantly lower than the controls. No differences were found at the college level. The results on the individual scales tend to follow the same pattern. If we had included the 10 subjects discarded to make age equivalent at the grammar school level, the interaction effect would have been even more pronounced.

The interaction effects are difficult to interpret. They may indicate that the way the

Table 3

Analysis of Variance Results in Comparisons of Mothers of Schizophrenics and Mothers of Normals[a]

Score or scale	Child's diagnosis	Mother's education	Interaction
Summary Scores			
A—Authoritarian-Control	—	15.41**	5.97**
B—Hostility-Rejection	—	1.89	—
C—Democratic Attitudes	1.39	—	—
Scales			
2—Fostering Dependency	—	5.80**	1.11
3—Seclusion of the Mother	—	8.80**	1.80
4—Breaking the Will	3.18	13.07**	3.98*
5—Martyrdom	—	12.21**	1.20
6—Fear of Harming the Baby	—	2.76	7.90**
7—Marital Conflict	—	3.99*	1.33
8—Strictness	6.31*	2.80	2.85
9—Irritability	—	1.44	—
10—Excluding Outside Influences	3.23	10.51**	3.71*
11—Deification	3.02	11.07**	4.57*
12—Suppression of Aggression	—	4.24*	2.02
13—Rejection of Homemaking Role	1.68	—	—
15—Approval of Activity	1.35	3.82	2.07
16—Avoidance of Communication	1.71	10.30**	1.01
17—Inconsiderateness of Husband	—	—	—
18—Suppression of Sex	—	11.75**	4.62*
19—Ascendance of the Mother	—	11.13**	6.12**
20—Intrusiveness	—	9.78**	2.69
22—Acceleration of Development	—	4.10*	4.31*
23—Dependency of the Mother	—	3.53*	3.62*

[a] *F* ratios of less than one are indicated by dashes.
* Significant at or below the .05 level.
** Significant at or below the .01 level.

Table 4

Comparisons of Means of Mothers of Schizophrenics and Mothers of Normals by Levels on Summary Score A

Level of education	Schizophrenics' mothers	Normals' mothers	t	p
Grammar school	242.09	214.91	1.88	.05–.10
High school	191.58	219.50	2.86	.01
College	172.86	155.86	.94	.3–.4

mothers of schizophrenics respond to the test depends on their familiarity with the cultural expectations for parental attitudes, which, in turn, may depend on their education. The samples at each level are not adequate to justify a conclusion. There is, however, a greater relationship between education and parental attitudes in mothers of schizophrenics than in the controls. The correlation between Factor A and education in the original 47 mothers of schizophrenics was − .632. In the original 47 controls it was only − .287. The difference between these two correlations was significant beyond the .05 level.

The hypothesis of a difference between parental attitudes of mothers of schizophrenics and other mothers was not confirmed. It would be difficult to disprove this hypothesis, since the possibility always remains that other instruments or techniques may reveal a difference. Toms (10) compared 30 mothers of male schizophrenics and 30 control mothers, using a variety of instruments including an interview, vocabulary test, the MMPI, and a short TAT. Unfortunately, there was a difference in education between the two groups which was not taken into account in the analyses. However, the general conclusion was that the hypotheses of personality differences between mothers of schizophrenics and other mothers were not confirmed from any of the techniques. One interesting result, which bears on our speculations about our interaction effects, was that the mothers of schizophrenics scored significantly higher on the two MMPI scales measuring defensiveness (L and K). These findings suggest that further research on the "schizophrenogenic mother" hypothesis might do better using tests which control for social desirability and response suppression.

Summary

It was hypothesized that mothers of schizophrenics have more severe (controlling and rejecting) parental attitudes than mothers of normals. Forty-two mothers of schizophrenic patients and 42 control mothers were compared on the Parental Attitude Research Instrument. The hypothesis was not supported, although significant interactions were found between the two groups and levels of education.

Received July 1, 1957.

References

1. Arieti, S. *Interpretation of schizophrenia.* New York: Robert Brunner, 1955.
2. Freeman, R. V., & Grayson, H. M. Maternal attitudes in schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **50**, 45–52.
3. Fromm-Reichmann, Frieda. Notes on development of treatment of schizophrenics by psychoanalytic therapy. *Psychiatry*, 1948, **11**, 263.
4. Mark, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 185–189.
5. McFarland, R. L. Personal communication, 1957.
6. Rosen, J. N. The treatment of schizophrenic psychoses by direct analytic therapy. *Psychiat. Quart.*, 1947, **21**, 3–37.
7. Schaefer, E. S., & Bell, R. Q. Development of the Parental Attitude Research Instrument. *Child Develpm.*, in press.
8. Shoben, E. J., Jr. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.*, 1949, **39**, 101–148.
9. Tietze, T. A study of mothers of schizophrenic patients. *Psychiatry*, 1949, **12**, 55–65.
10. Toms, E. C. Personality characteristics of mothers of schizophrenic veterans. Unpublished doctoral dissertation, Univer. of Minnesota, 1955.
11. Zuckerman, M., Ribback, Beatrice B., Monashkin, I., & Norton, J. A. Normative data and factor analysis on the Parental Attitude Research Instrument. *J. consult. Psychol.*, 1958, **22**, 165–171.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Representative Elements in Role Construct Technique

Spiro B. Mitsos
Evansville State Hospital

According to the psychology of personal constructs as set forth by Kelly (2), each individual has a conceptual system (an organized set of personal constructs) with which he gives structure to the world around him. Basic to the theory is the constant attempt by each individual to anticipate events in his surrounds. Hence, man is thought of in somewhat the same light as the scientist in his continual process of prediction and validation. Just as the scientist begins with a theory and makes his predictions on the basis of the concepts from that theory, man has a theory of life composed of bipolar personal constructs. His predictions are made along the lines of these personal constructs.

In applying this theoretical framework, Kelly and his associates have developed a technique designed to elicit the construct framework of any given individual. This technique is the Role Construct Repertory Test (RCRT).

With the RCRT, the subject (*S*) thinks about certain people he knows—people meaningful to him and who represent a variety of types. These people are referred to as "elements" and are determined by a role-title list. The *S* assigns people he knows to fulfill the role titles. Examples are as follows:

1. Write your own name in the first blank here.
8. Your closest present friend of the same sex as yourself.
17. The most successful person whom you know personally.

The *S* is then given these elements three at a time and asked to think about the three people and decide some important way in which two are alike and which sets them off from the third person. The contrasts are listed. Kelly assumes they are important constructs in the individual's personal construct system.

Certain other assumptions underlie the RCRT. One of these is the assumption of the "representativeness of the elements." "If the test is to indicate how the subject develops his role in the light of his understanding of other people, it is necessary that the other people appearing as elements in the test be sufficiently representative of all the people with whom the subject must relate his self-construed role" (2, p. 230). On the basis of this assumption, Kelly constructed a list of role titles; i.e., self, mother, father, ex-flame, pal, authorities, etc.

This assumption is the focus of concern in this paper. Kelly assumes representative elements (representative role titles) are necessary to elicit the most crucial personal constructs. The present research began with the simple question: "What would happen if the representative elements assumption were violated—and why?"

The problem was further limited to the question of consistency of construct labels with and without the fulfillment of the above assumption. Hunt (1) has previously investigated certain aspects of consistency with the RCRT. He examined *S*s on two occasions with a one-week time interval. Essentially different role-title lists were used each time. However, the two lists were equated as to sex, relatives, authority figures, etc. With his technique of determining equivalence of constructs, he found the average percentage of agreement in constructs used on the two protocols to be 69% with a standard deviation of 6%.

In Hunt's research, the assumption of representative elements was fulfilled in both forms of the test. In the present research, we are concerned with the question of whether or not similar results will occur if the assumption is violated in one form. The predictions were as follows: Construct labels elicited by the RCRT will be consistent

through time when a representative role-title list is used. Consistency will also occur when the role-title list is nonrepresentative, in this case, homogeneous.

Procedure

The *Ss* for this investigation included 20 psychiatric aides randomly divided into two equal groups. One of these groups is referred to as the "representative elements group (REG)"; the other, the "nonrepresentative elements group (NREG)." Constructs were elicited from the *Ss* in the REG in the same manner as the Role Construct Repertory Test as described by Kelly. (In the present form there were 19 role titles and 22 sorts of 3 elements each.) All *Ss* were re-examined with the same technique after a three-month interval. The *Ss* in the NREG were treated in the same manner with a singular exception—the role-title list was not used. In its place, each *S* was merely asked to list the names of 19 personal friends.

According to the hypothesis, we would expect a significant number of constructs elicited on the first administration to be repeated on the second administration. Such would apply to both groups. Each construct for each *S* on the first administration was evaluated individually. If either pole of a given construct were repeated verbatim anywhere in the entire repertory elicited on the second administration, this was considered as repetition. Binomial expansion was utilized to establish 5% significance levels for number of constructs repeated and number of *Ss* repeating a significant number of constructs.

Results

All protocols were evaluated by comparing both protocols for each *S*. If either or both poles of any given construct on the first administration were repeated verbatim on the second administration, a replication was scored. Thus the potential range of scores was 0–22. Applying the binomial expansion, 16 of 22 replications would be significant at the .05 level.

In the representative elements group, nine of the ten *Ss* repeated significant numbers of constructs ($p < .02$). In the nonrepresentative elements group, two *Ss* did so ($p > .20$).

Table 1

Distribution of Replicated Constructs and Significant Subjects

	REG	NREG	χ^2	p
Constructs Repeated	170	122		
Constructs Not Repeated	50	76		
Significant Subjects	9	2	6.36	<.02
Nonsignificant Subjects	1	7		

(One *S* in the latter group had to be dropped because he was not available for re-examination.) The results seem to indicate a direct relationship between representative role titles, on which to base constructs, and consistency of construct labels.

Chi square was computed for the two groups comparing the number of *Ss* repeating significant numbers of constructs with the number repeating insignificant numbers. These data are presented in Table 1.

Certain questions arose at this point of the research. Perhaps the most obvious was the possibility that these results are the result of the role-title list yielding the same people on each administration, while a different group of friends was used each time. Another question is that of whether or not people, when responding to such a test, are responding to the types of people described on the role-title list rather than or in addition to the particular people involved? Lastly, are different types of constructs elicited when the elements are heterogeneous as opposed to homogeneous?

In an attempt to clarify the results, the *Ss* in the NREG were re-examined after another three-month interval. On this administration, *Ss* used the identical list of 19 friends they had produced on the second administration. These friends were listed in identical order, and the *Ss* were given identical sorts to elicit constructs. In comparing the second to the third administration, three *Ss* repeated significant numbers of constructs ($p > .20$)—substantially the same results as the previous administration.

Discussion

It would seem apparent, on the basis of the above, that fulfillment of the assumption of

the representativeness of the elements is a matter of importance, at least in terms of consistency. Construct labels were consistent through time when the assumption was fulfilled and were inconsistent through time when the assumption was violated.

It would appear, therefore, that conceptual consistency on the RCRT, at least insofar as it applies to the construct labels, must be in some way a function of the role-title list and the types of constructs elicited thereby.

The reasons can only be conjectured. The most likely source of explanation is Kelly's theory itself. The fundamental postulate of his system is stated as follows: "A person's processes are psychologically channelized by the way in which he anticipates events" (2, p. 46). One of the corrolaries of this postulate, the "Organization Corrolary," states: "Each person characteristically evolves, for his convenience in anticipating events, a construction system embracing ordinal relationships between constructs" (2, p. 56).

According to this corrolary, personal constructs within any individual's system are arranged in a hierarchical structure. "Within a construction system there may be many levels of ordinal relationships, with some constructs subsuming others and those, in turn, subsuming still others" (2, p. 57). Subsuming constructs are termed superordinal and subsumed constructs subordinal.

Relating these corrolaries to the present data, it would seem plausible that, when thinking of the people represented on the role title list, *Ss* use personal constructs capable of describing not only specific individuals, but also describing similarities and differences between groups or types of people, i.e., family, authorities, peers, etc. Description in this case refers to the process of labeling dimensions on which one could anticipate the behavior of these individuals and groups.

In the present research, the "nonrepresentative elements group" described only a group of friends. These *Ss* were not concerned with describing groups; their sample of people was fairly homogeneous. They were concerned, as it were, with only one possible subgroup of the heterogeneous group used by the "representative elements group." The construct of "friend versus nonfriend" might well be one of the group of dimensions used with the heterogeneous group of people.

In other words, it seems plausible that consistency occurs when the task requires the use of generally superordinate constructs. More specifically, it might be hypothesized that the number of superordinate constructs produced in a concept formation situation is a function of the heterogeneity of the objects to be construed. It may well be that each individual has an infinite number of subordinate constructs and hence a great deal of leeway in construing within the context of a superordinate construct; in this case, the context of friend versus nonfriend. Such an explanation would seem to be consistent with Kelly's fundamental postulate and the "Organization Corrolary."

Summary

The purpose of this study was essentially that of testing the differential effects of heterogeneity and homogeneity of role titles on the Role Construct Repertory Test. Two groups of subjects were administered the test. One group utilized a heterogeneous role-title list, the other a homogeneous group of elements. Significant differences were found in terms of the consistency of construct labels. In the group using the role-title list, a significant ($p < .02$) number of subjects demonstrated consistency of construct labels through time. In the group using homogeneous elements, only 20 per cent ($p > .20$) of the subjects demonstrated such consistency.

It would appear most likely that such results are a function of the heterogeneity of the role-title list. When presented with a heterogeneous group of stimulus objects on which to perform sorts and attach construct labels, individuals tend to produce more superordinate constructs, whereas more subordinate constructs are produced when the construed objects are homogeneous.

Received July 18, 1957.

References

1. Hunt, David E. Studies in role concept repertory: Conceptual consistency. Unpublished master's thesis, Ohio State Univer., 1951.
2. Kelly, George A. *The psychology of personal constructs.* New York: Norton, 1955.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Taylor A Scale and Clinical Measures of Anxiety[1,2]

Carl G. Lauterbach
Madigan Army Hospital

This study was designed to determine the relationships of the Taylor A Scale to clinical criteria of manifest anxiety, and to compare these relationships with those between Welsh's Anxiety Index (AI) and the same clinical criteria.

The *S*s were 44 male members of the U. S. Armed Forces referred to a military hospital for psychiatric consultation.

Three psychologists and a psychiatrist independently rated each *S* on a combined check list and rating scale of anxiety. The psychiatrist completed his ratings after his initial diagnostic interview. Then the psychologists simultaneously performed their ratings while one of them also administered the Wechsler-Bellevue Intelligence Scale to the *S*. Finally, the MMPI was administered, which provided both the A scale and AI variables.

The intercorrelations among the psychologists in their anxiety ratings ranged between .78 and .80. However, none of their ratings showed a significant relationship to the psychiatrist's ratings. Thus, a composite of the three psychologists' ratings was treated as a separate criterion of anxiety from the psychiatrist's ratings. An item analysis of the check list of anxiety signs resulted in four signs being selected, assigned weights, and summed for each *S* to form a third criterion of anxiety, referred to as the composite weighted check list of the psychologists. A psychiatric diagnosis formed the fourth criterion of anxiety.

The A scale correlated .44 ($p < .01$) with the composite rating of the psychologists and .38 ($p < .05$) with the composite weighted check list. The AI correlated .36 ($p < .05$) with the composite rating of the psychologists. Neither the A scale nor the AI correlated significantly with the psychiatric criteria of anxiety.

The two anxiety test measures and the four criteria of anxiety were correlated with the three validity scales of the MMPI and with the $F - K$ dissimulation index. The A scale related significantly to the *F*, *K*, and $F - K$ scales, while there was a lack of significant relationships, in general, between the validity scales and the criteria of anxiety. In the case of the *K* scale, it appears that *K* could be used as a variable to suppress the error in the A scale to improve the latter's relation to the criterion measures.

Brief Report.
Received March 15, 1958.

[1] This study is based upon a dissertation submitted for the Ph.D. degree, University of Iowa, 1952. The writer wishes to express his gratitude to Grant W. Dahlstrom, Committee Chairman, and the several staff members of the Neuropsychiatric Service, Fitzsimmons Army Hospital, who so freely participated in obtaining the data.

[2] An extended report of this study may be obtained without charge from Carl G. Lauterbach, 243 South 80th St., Tacoma, Wash., or for a fee from the American Documentation Institute. Order Document No. 5580, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Experience of Clinicians, and Speech Characteristics and Statements of Clients as Variables in Clinical Judgment[1]

Austin E. Grigg[2]
State University of Iowa

Whether he is reading a case history, conducting a psychological test, or interviewing a client, the clinical psychologist has as a major objective the formulation of certain kinds of judgments about his client. On the basis of these judgments, the psychologist develops his plans for the management of the case or for the kinds of recommendations he will make in a professional report. Very little is known, however, about the variables that are operating in making these judgments.

There appear to be some empirical findings to suggest that clinical experience, per se, does not insure more accurate clinical judgments about clients, as shown by Luft (9, 10), Cline (2), and Soskin (11). Some studies show that the method by which clinicians acquire information about the client may not have to be the traditional interview procedure in order for accurate judgments to be made, as, for example, the Borke and Fiske study (1) finding that clinicians were able to make as accurate predictions about neurotic patients after reading interview statements as they could make after face-to-face interactions with these patients, and Luft's report that readers of interviews did as well as listeners in predicting client responses to objective questionnaires. Nor is it clear today just how much information a clinician needs about a client in order to make valid judgments, or what kinds of data he needs. Stern, Stein, and Bloom (12), for example, found that increased observational and test data tended to be associated with more accurate judgments about individuals, but Gage (3) reported a lowering of judgmental accuracy with increased observational data and Kostlan (7) found that unless a social history was included, increased clinical testing data did not make for better predictions.

In an effort to learn more about the variables in an interview situation that may play a role in influencing the judgments of clinicians about their clients, three variables were manipulated in the present study: the voice of the interviewee, the content of the interview statements, and the training and experience of the clinician.

Procedure

The first counseling interviews of three male clients with personal problems who reported to a university counseling service were tape recorded. During this first visit, the clients indicated their appraisal of themselves by making *Q* sorts with Deck 10 from Hilden's *Random Sets of Personal Concepts* (6), by answering Gough's Adjective Check List (4), and a 25-item Self Report Questionnaire adapted from the Guilford-Martin Inventory of Factors GAMIN (5). After repeating these measures one week later, each of the clients was included in the study since the reliabilities of the *Q* sorts were .80 or higher, and the percentage of overlap of items on the other two measures exceeded 85 per cent.

From each of the clients' interviews, a ten-minute portion of the tape recording was selected on the basis of giving a general state-

[1] This paper is based upon a doctoral dissertation completed at the State University of Iowa, August, 1957. The writer acknowledges the advice and suggestions of Leonard D. Goodstein, under whose direction this study was done. Gratitude is also expressed to Leonard Feldt for his consultation about statistical processing of the data.

[2] Now at the Department of Psychology and the Testing and Guidance Bureau, The University of Texas.

ment of the client's complaints without stress on specific etiological material, and an exact typescript was prepared. Then, an experienced actor re-enacted the interview selection with the original counselor in whatever manner the actor felt the script indicated, but without the actor hearing any recordings of the actual client. The same male actor was used to re-enact each script.

Thus, for each of the three clients, any one of three sets of data could be supplied to a given group of clinical judges: (*a*) the actual tape-recorded selection with the client, (*b*) the tape-recorded selection as re-enacted by the actor, or (*c*) an exact typescript of the interview selection. Also, since there were selections from interviews of three clients, it was possible to vary the content of the interviews.

Twenty-four male clinical psychologists with the Ph.D. degree and affiliation with the American Psychological Association were used as the Ph.D. judges. Twenty-four male trainees in clinical psychology constituted the Trainee judges and included either graduate students with at least one year of practicum experience, or clinical trainees in a hospital who had completed at least one year of graduate work in psychology. A group of Naive judges, 24 male undergraduate students from the introductory psychology course, were obtained from Coe College. The Ph.D. and Trainee judges were obtained from three Veterans Administration hospitals: Downey, Illinois, Iowa City, and Knoxville, Iowa; from the Elgin (Illinois) State Hospital, and from the Psychopathic Hospital and the Department of Psychology of the State University of Iowa.

Groups of judges were given various combinations of the voice and content conditions and after each presentation were asked to predict (actually, postdict) how the client responded to the three personality tests. The accuracy of any judge's predictions was measured by the number of agreements between these predictions and the responses of the clients, called "Client Key." Later, although they were instructed to predict the responses of the clients and not the impressions of other psychologists about the clients, the judges' predictions were scored by an "Expert Key" made from a consensus of three psychologists as to what the clients' responses "should be" after these psychologists had studied, independently, the complete case folder of each client.

Each of the three experience levels of judges was subdivided into three subgroups of eight judges, and the sequence of procedure was so arranged that any one subgroup heard a different voice and content combination than either of the other two subgroups. Each subgroup, but with different voice-content combinations, had the following three conditions: (*a*) a client speak his own interview content, (*b*) an actor re-enact a different one of the three interviews, and (*c*) an exact typescript of the third and different interview selection. The order of presentation was confounded with sequence subgroups, so that each of the above conditions occurred in each possible order of presentation as first, second, or last. By using a Lindquist Type IV analysis of variance design (8, pp. 285–288), the voice and the content conditions were "within" effects, and the experience levels were "between" effects.

Results

Table 1 presents the means and the standard deviations of the three variables—experience, voice, and content—as measured by the Client Key. A summary table of *F* ratios from

Table 1

Means, Standard Deviations of Prediction of Client Responses with Variables Manipulated in the Study

	Q Sort		SR		Gough ACL[a]	
	Mean	Sigma	Mean	Sigma	Mean	Sigma
Experience:						
Naive	.300	.063	14.46	1.51	37.29	18.68
Trainee	.368	.063	15.85	1.22	42.89	13.90
Ph.D.	.376	.066	16.43	1.17	42.39	20.87
Voice:						
Client	.344	.107	16.39	2.22	39.67	16.73
Actor	.333	.124	14.60	2.80	38.81	14.99
Script	.367	.108	15.75	2.18	44.10	16.49
Content:						
#1	.328	.112	15.25	2.44	51.72	14.92
#2	.366	.128	14.83	2.93	33.53	12.41
#3	.350	.099	16.65	1.68	37.32	15.23

[a] To reduce 300-item Gough ACL scores to smaller numbers, 190 was subtracted from total agreements between judge and client responses.

Table 2

Summary Table Showing *F* Values for the Three Measures

Key	*Q* Sort		Self Report		Gough ACL	
	Client	Expert	Client	Expert	Client	Expert
Between Effects						
Experience	9.55**	11.15**	13.76**	3.95*	2.47	0.01
Voice-Content (b)	1.02	3.39*	1.15	0.67	6.98**	3.68*
Voice-Cont.-Exp. (b)	0.53	2.61	1.08	0.88	2.09	0.31
Within Effects						
Voice	1.93	12.41**	15.02**	12.47**	5.32**	2.24
Content	2.35	54.62**	16.55**	22.50**	60.83**	16.93**
Voice-Content (w)	1.87	1.71	8.69**	10.12**	9.85**	0.72
Voice-Experience	0.56	1.74	0.79	0.71	0.38	0.98
Content-Experience	0.26	1.04	0.48	1.84	2.35	2.64
Voice-Cont.-Exp. (w)	1.68	0.36	1.55	0.04	0.68	0.51

* Significant beyond .05 level of confidence.
** Significant beyond .01 level of confidence.

the analysis of variance is presented in Table 2.

With the *Q* sorts, it can be observed that the two trained groups of judges were more accurate in their predictions than the naive group, and this difference was shown by further analysis to be statistically significant ($p < .01$), but the difference between the accuracy of predictions by Trainee judges and Ph.D. judges is not statistically significant. The differences between the three voice conditions are not significant with the *Q* sorts, although the actor's voice tended to lower the accuracy of the predictions about the client's responses. Also, with the *Q* sorts, when different interview contents were used, there were no statistically significant differences in the accuracy of the predictions.

With the Self Report Questionnaire (SR), both groups of trained judges were significantly more accurate in predicting the clients' responses than were the Naive judges ($p < .01$), but again the difference between the Trainee judges and the Ph.D. judges was not significant. With SR, the three voice conditions and also the three different contents of the interviews influenced the accuracy of the judges significantly, as shown in Table 2. Because there was a Voice-by-Content interaction, the effects of voice conditions were compared over each content separately and it was found that the client's voice produced the most accurate predictions for Content 1, but no significant voice effects occurred with the other content conditions with SR.

With the Gough Adjective Check List, although the two trained groups again appear to be superior to the Naive judges, statistical analysis summarized in Table 2 indicated that the differences are not statistically significant. Again, the actor's voice was associated with the poorest predictive accuracy. Content 1 was associated with accuracy scores strikingly superior to those obtained when Content 2 or Content 3 occurred ($p < .01$). Because with Gough ACL a Voice-by-Content interaction occurred, an analysis similar to that described above with SR data was performed, and it was observed that the client's voice was associated with significantly *lower* accuracy scores with Content 3, but this did not occur with the other contents.

With the Client Key, the three measuring instruments used—*Q* sorts, SR, and Gough ACL—reacted differently to the variables of the experiment (Table 1). With the *Q* sorts and the Gough ACL, the script gave the most favorable scores for the voice conditions. With each of the three personality measures, the actor's voice lowered the accuracy of the predictions. In regard to the content of the interviews, the three personality measures reacted differently: Content 1 gave strikingly superior scores over other contents with the Gough ACL, but was the poorest content condition with *Q* sorts. Content 2 was associated

with the best scores on the *Q* sorts, but was the poorest content condition with the other two measures. With SR, Content 3 was associated with the highest predictive accuracy.

Table 3 presents the means and the standard deviations of the three variables as scored by the Expert Key. The three clinicians who composed the key, working independently, answered the three personality measures as they believed the client "should have responded" after they had studied all pertinent case history and interview reports as well as the actual recorded selections heard by the judges in the experiment. The average *Q* sort of the three judges and those SR and Gough ACL items on which two or more of the three "experts" agreed, constituted the Expert Key. As may be seen, the trends in the data are similar to those of the Client Key (compare Table 1 and Table 3), although now the accuracy scores are considerably higher. The major discrepancy between the results based on the two keys occurs in the *Q*-sort data. There, significant voice and content effects are revealed with the Expert Key, as shown in Table 2. In addition, the "between" component of the Voice-by-Content interaction was significant, and the latter effect also occurred with the Gough ACL data. Such a "between" effect may be associated with the disproportionate number of State University of Iowa judges in the three sequence subgroups. Subgroup I contained twelve such judges, whereas Subgroup II contained seven, and Subgroup III contained none. While this effect did not influence the predictions of how the clients responded, it apparently exerted an influence when the judges were scored by an Expert Key devised by three State University of Iowa psychologists.

Table 3

Means, Standard Deviations with the Variables Manipulated in the Study (Expert Key)

	Q Sort		SR		Gough ACL	
	Mean	Sigma	Mean	Sigma	Mean	Sigma
Experience:						
Naive	.503	.092	16.62	1.65	45.42	18.63
Trainee	.599	.075	17.96	1.44	45.54	13.59
Ph.D.	.575	.062	17.57	1.72	45.89	13.78
Voice:						
Client	.588	.145	18.31	3.17	45.67	18.72
Actor	.506	.167	15.97	3.99	43.65	18.71
Script	.584	.313	17.88	2.79	47.53	19.87
Content:						
#1	.623	.151	17.69	3.64	51.57	18.65
#2	.447	.141	15.58	3.35	41.22	18.63
#3	.608	.096	18.88	2.62	44.06	18.74

Because of the relatively large differences between the average scores under the Client and Expert keys, it was considered expedient to test the significance of these differences by the simple nonparametric sign test (13, pp. 430–435). The results of this test clearly verified the observation that the judges within each group show greater agreement with the Experts' predicted responses for the clients than with the responses of the clients themselves.

Discussion

From the results of this study, it appears that judges with one or more years of clinical experience are able to make more accurate predictions about clients' responses to certain personality measures than can be predicted by untrained judges. But among judges who have had at least one year of training, increased experience does not appear to be associated with a significant increase in the accuracy of predictions. There is the problem of whether these results imply that whatever skill is added to the individual's ability to make judgments along these lines is acquired rapidly, during the early period of graduate training, and that increments thereafter are too gradual to make for significant experience level differences. The results could be interpreted also to suggest that selection may be the important factor here: that certain kinds of people, better than average in ability to judge the responses of people who are upset, enter clinical psychology and that training to the level of the Ph.D. does not increase their skills along these lines. Another possibility is that perhaps the clients used in this study placed the more experienced Ph.D. judges at a disadvantage, since they work mostly in hospital settings, and the clients used in the study were outpatients of a university counseling service.

In the present study, unlike the Luft studies where clinicians were no more accurate than nonclinicians, brief interview selections were presented to the judges. It could be that the

trainee and the experienced clinician are able to formulate better personality appraisals on the basis of little data than are naive judges. Luft's findings with one-hour interviews that experienced clinicians are no more accurate than inexperienced judges suggests the possibility that as the amount of interview material available as a basis for judgment increases, the superiority of the experienced judges over naive judges may diminish. Or, there may be differences due to age differences, since in the present study, the Naive judges were undergraduates. But these speculations must await empirical testing.

While it is true that, for all three personality measures employed in this study, a discrepant (the actor's) voice lowered the accuracy of clinical judgment, reading the interview selection allowed more accurate predictions of *Q* sorts and Gough ACL responses than did listening to the client. This corresponds to Luft's report that student readers did as well as student listeners in predicting responses to objective questionnaires, and seems related to Borke and Fiske's finding that reading interview statements was associated with as accurate predictions as occurred after face-to-face interaction with the client.

In this study, no significant Voice-by-Experience or Content-by-Experience interactions occurred. The experience groups were more accurate than the Naive judges over all conditions studied, and not just one content or one voice.

Conclusions

1. Two groups of judges with clinical training and experience, the Trainees and the Ph.D. judges, predicted client responses more accurately than the Naive judges.

2. Between the experienced groups, however, there occurred no statistically significant differences in the accuracy of prediction of client responses.

3. A voice discrepant from the original speaker of the interview content lowered the accuracy of the judgments for all groups of judges.

4. Different interview contents appeared to be related to differences in accuracy of judgments about the client, but the instrument being used appeared to be an important variable here.

5. The script condition, with total absence of voice cues, was associated with the most accurate predictions with two of the three measuring instruments used.

6. There was no statistically significant interaction between experience and voice, or between experience and content.

7. The judges, regardless of experience level, were in greater agreement with what a consensus of "experts" predicted about client responses than in predicting the actual responses made by the clients themselves.

Received July 17, 1957.

References

1. Borke, Helene, & Fiske, D. W. Factors influencing the prediction of behavior from a diagnostic interview. *J. consult. Psychol.*, 1957, **21**, 78–80.
2. Cline, V. B. Ability to judge personality assessed with a stress interview and sound film technique. *J. abnorm. soc. Psychol.*, 1955, **50**, 183–187.
3. Gage, N. L. Explorations in the understanding of others. *Educ. psychol. Measmt*, 1953, **13**, 14–26.
4. Gough, H. G. *Reference handbook for the Gough Adjective Check List.* Berkeley, Calif.: Univer. Calif. Inst. Pers. Assessment and Res., 1955.
5. Guilford, J. P., & Martin, H. G. The Guilford-Martin Inventory of Factors GAMIN. Beverly Hills, Calif.: Sheridan Supply, 1943.
6. Hilden, A. H. *Manual for Q-Sort and random sets of personal concepts.* Webster Groves, Mo.: Author, 1954.
7. Kostlan, A. A method for the empirical study of psychodiagnosis. *J. consult. Psychol.*, 1954, **18**, 83–88.
8. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1956.
9. Luft, J. Implicit hypotheses and clinical predictions. *J. abnorm. soc. Psychol.*, 1950, **45**, 756–759.
10. Luft, J. Differences in prediction based on hearing versus reading verbatim clinical interviews. *J. consult. Psychol.*, 1951, **15**, 115–119.
11. Soskin, W. F. Bias in postdiction from projective tests. *J. abnorm. soc. Psychol.*, 1954, **49**, 69–74.
12. Stern, G., Stein, M., & Bloom, B. *Methods in personality assessment.* Glencoe, Ill.: Free Press, 1956.
13. Walker, Helen M., & Lev, J. *Statistical inference.* New York: Holt, 1953.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

A Note on the Reliability of the Leiter International Performance Scale 1948 Revision[1]

Heber C. Sharp
Utah State University

Since the Leiter International Performance Scale (LIPS) was developed, most of the reported research has been designed to explore the validity of the LIPS by comparing it with other measures of intelligence. Reliability studies seem to have been neglected. Leiter (1) states that two types of reliability data are available. The type reported most frequently compares the standard deviation of the LIPS results with those of the accepted instrument. Leiter reports a split-half reliability coefficient of .89 (1, p. 30). Tate (2), following Thorndike's outline (3, pp. 93–95) of Holt's analysis procedure to estimate reliability, reports a coefficient of .75. The following data are concerned with a test-retest reliability coefficient.

Procedure. The LIPS was administered to 48 retarded school children in September or October and again in April or May with at least a six-month interval between tests. The *Ss* ranged in age from 8 years and no months to 16 years and 5 months. 10 *Ss* were attending a special school for seriously retarded children, 21 were attending special classes for retarded children, and 17 were attending classes with other children in the regular schools.

Results. Statistical data revealed for the initial test an M of 65.1, SD 13.67, and the range of 38–88. The retest M was 67.77, SD 13.73, and the range 41–95. The reliability coefficient was found to be .91. A t test to determine the significance of the difference between the means yielded a t of 3.15 ($p < .01$).

[1] This study was supported in part through a grant from Utah State University research program.

Discussion. The retest method may introduce a spurious factor due to carry-over from one administration to the other. *Ss* frequently mentioned memory for one or more test items. This might account for the difference found between the means. Only 6% of the *Ss* had the same score on both test and retest, 27% earned lower scores on the retest, and 67% showed increases from 1 to 14 IQ points on the retest. For this retarded group, some items seemed to be improperly placed in the scale. Further empirical work concerned with this problem is suggested.

The reliability coefficient of .91 indicates both stability and consistency of the LIPS as a measuring instrument for *Ss* at the lower end of the intelligence distribution. The similarity of the variance measures supports further the reliability of the LIPS. Since the *Ss* were drawn from a restricted range of ability, thus limiting the spread of scores, the high coefficient does not imply necessarily reliability for all types of subjects.

Received August 12, 1957.

References

1. Leiter, R. G. *Part I of the manual for the 1948 revision of the Leiter International Performance Scale.* Chicago: C. H. Stoelting, 1952.
2. Tate, Mariam T. The influence of cultural factors on the Leiter International Performance Scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 497–501.
3. Thorndike, R. L. *Personnel selection test and measurement techniques.* New York: Wiley, 1949.

HOMOGENEOUS GROUPING

The size of The Devereux Schools makes possible homogeneous grouping of students. Boys, girls, and young adults of similar aptitude, maturity, and emotional problems are placed in one of the nineteen separate school units in Pennsylvania—or in similar groups in California. Within the semi-autonomous group the specially trained staff sets the pace to suit ability—not too fast, not too slow. Methods are adapted to need. In this environment students tend to overcome diffidence, gain confidence, and make real progress.

And yet the boys and girls in these small, residential groups have the advantages that only a sizeable organization brings. The Devereux Foundation numbers on its professional staff psychiatrists, psychologists, general practitioners, and educational personnel. This enables the slow-learning and emotionally disturbed students to have the advantage of the multidisciplined approach to their problems.

Professional inquiries should be addressed to John M. Barclay, Director of Development, Devereux Schools, Devon, Pennsylvania; western residents address Keith A. Seaton, Registrar, Devereux Schools in California, Santa Barbara, California.

THE DEVEREUX FOUNDATION

A nonprofit organization — *Founded 1912*

Santa Barbara, California — *Devon, Pennsylvania*

SCHOOLS
COMMUNITIES
CAMPS
TRAINING
RESEARCH

HELENA T. DEVEREUX, Administrative Consultant

EDWARD L. FRENCH, Ph.D., Director

Professional
Associate Directors
Charles M. Campbell, Jr., M.D.
Michael B. Dunn, Ph.D.
Fred E. Henry, S.T.D.
J. Clifford Sent, M.D.

PSYCHOLOGICAL TEST SPECIALISTS

Exclusive publishers and distributors for selected psychological tests

● ● TESTS NOW AVAILABLE ● ●

KAHN TEST OF SYMBOL ARRANGEMENT

Highly versatile projective test using symbolic plastic objects which are arranged and sorted. Validly discriminates among normals, neurotics, brain-injured, and psychotics, and indicates accompanying personality dynamics.

$25.00 for set of materials and manuals

FULL-RANGE PICTURE VOCABULARY TEST

Brief test of intelligence in two forms using the same 16 plates. Words are defined by choosing pictures. Valuable for intelligence screening, or with the physically handicapped.

$7.50 for set of materials

PROVERBS TEST

Group and individual forms determine similarity of thought processes to those of various normal and clinical groups. Highly sensitive indicator of psychotic processes.

$4.50 for kit

VOCATIONAL APPERCEPTION TEST—COLLEGE FORM

Separate male and female series of plates depicting college-related occupational activities, accompanied by brief explanations of use and interpretation as a projective test.

$7.25 for set of materials

● TESTS TO BE PUBLISHED IN THE IMMEDIATE FUTURE ●

EMOTIONAL PROJECTION TEST

Emotional states are attributed to persons shown in 30 photographs. Extensive research has shown that responses to this test are extremely sensitive indicators of needs and feelings and can be effectively analyzed to determine personality structure.

FAMOUS SAYINGS

Agreement with proverbs, aphorisms, and folk sayings is analyzed to determine personality structure. Of particular use in personnel selection.

GEIST PICTURE INTEREST INVENTORY

Provides for expression of vocational interests by choices among alternative pictured activities. Qualitative and quantitative interpretations are possible. A relatively non-verbal test with great potential both for children and for adults.

ID-EGO-SUPEREGO TEST

Subjects give independent estimates of personality components from a modified psychoanalytic point of view.

QUICK PICTURE VOCABULARY TEST

Brief, three forms based on one test plate; gives a good estimate of intellectual ability in 2-3 minutes. Can be used independently or as third form of FRPV.

● ● TESTS IN PREPARATION ● ●

Vocational Apperception Test—General Form; Electron Micrograph Apperception Test; Bi Scale for Children; Group Personality Projective Test; Projective Movement Sequences.

To order any of the tests now available, or to obtain information about those to be published, write

Psychological Test Specialists
Box 1441
Missoula, Montana